Semiclassical Analysis of Loop Quantum Gravity

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Herr Dipl.-Phys. Florian Conrady
geboren am 23.12.1976 in Ulm

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Thomas Buckhout, Ph.D.

Gutachter:

1. Prof. Carlo Rovelli
2. Prof. Dr. Thomas Thiemann
3. Prof. Dr. Hermann Nicolai

Tag der mündlichen Prüfung: 17. Oktober 2005

Abstract

In this Ph.D. thesis, we explore and develop new methods that should help in determining an effective semiclassical description of canonical loop quantum gravity and spin foam gravity. A brief introduction to loop quantum gravity is followed by three research papers that present the results of the Ph.D. project.

In the first article, we deal with the problem of time and a new proposal for implementing proper time as boundary conditions in a sum over histories: we investigate a concrete realization of this formalism for free scalar field theory. In the second article, we translate semiclassical states of linearized gravity into states of loop quantum gravity. The properties of the latter indicate how semiclassicality manifests itself in the loop framework, and how this may be exploited for doing semiclassical expansions. In the third part, we propose a new formulation of spin foam models that is fully triangulation- and background-independent: by means of a symmetry condition, we identify spin foam models whose triangulation-dependence can be naturally removed.

Keywords:
Loop quantum gravity, Spin foams, Lattice gauge theory, Classical and semiclassical methods

Zusammenfassung

In dieser Dissertation untersuchen und entwickeln wir neue Methoden, die dabei helfen sollen eine effektive semiklassische Beschreibung der kanonischen Loop-Quantengravitation und der Spinfoam-Gravitation zu bestimmen. Einer kurzen Einführung in die Loop-Quantengravitation folgen drei Forschungsartikel, die die Resultate der Doktorarbeit präsentieren.

Im ersten Artikel behandeln wir das Problem der Zeit und einen neuen Vorschlag zur Implementierung von Eigenzeit durch Randbedingungen an Pfadintegrale: wir untersuchen eine konkrete Realisierung dieses Formalismus für die freie Skalarfeldtheorie. Im zweiten Artikel übersetzen wir semiklassische Zustände der linearisierten Gravitation in Zustände der Loop-Quantengravitation. Deren Eigenschaften deuten an, wie sich Semiklassizität im Loop-Formalismus manifestiert, and wie man dies benützen könnte, um semiklassische Entwicklungen herzuleiten. Im dritten Teil schlagen wir eine neue Formulierung von Spinfoam-Modellen vor, die vollständig Triangulierungs- und Hintergrund-unabhängig ist: mit Hilfe einer Symmetrie-Bedingung identifizieren wir Spinfoam-Modelle, deren Triangulierungs-Abhängigkeit auf natürliche Weise entfernt werden kann.

Schlagwörter:
Schleifen-Quantengravitation, Spin-Schäume, Gittereichfeldtheorie, Klassische und semiklassische Methoden

Contents

List of Figures

Chapter 1

Introduction

One of the greatest challenges of modern theoretical physics is the unification of general relativity and quantum theory. In spite of considerable efforts over decades, one has not yet succeeded in formulating a theory that would describe the observed physics of gravitation and microscopic matter in a coherent manner. In the quest for such a formulation, the two most developed approaches are currently superstring theory and loop quantum gravity.

The term "loop quantum gravity" actually subsumes two directions of research, which are closely related, but not always equivalent: we refer to them as *canonical loop quantum gravity* and *spin foam gravity* respectively.

On the conceptual level, loop quantum gravity (LQG) has three principal features:

1. its formulation is based on the concept of connection,

2. it is manifestly background-independent,

3. and, in some cases, the quantization procedure could imply an inherent UV cutoff.

The use of connections can be motivated by the fact that the three forces of the standard model are all described by gauge theories, and even classical gravity can be cast in a connection formulation. The principle of background-independence incorporates a lesson we learned from classical gravity, namely, that there is no fixed geometry of spacetime on which the rest of physics happens; instead, the geometry is itself a dynamical variable. The third point refers to the fact that in certain models of LQG, the gauge group is compact and leads to a minimal value of length, area and volume eigenvalues.

Canonical LQG results from a quantization of a Hamiltonian formulation of classical GR. In 4 dimensions, the original approach was based on the Ashtekar form of gravity, which involves complex connections. Nowadays, one mostly uses a real connection formulation with the so-called Ashtekar-Barbero variables. Wilson loops of the connection play a central role in the quantization, and that motivated the name "loop quantum gravity".

Spin foam gravity was developed later, and can be seen as an attempt to use similar techniques in a manifestly covariant manner. By that we mean that the viewpoint is shifted from a state-evolution to a sum-over-histories picture. There are different means to construct the models of spin foam gravity. In this thesis, we emphasize a viewpoint where the spin foam models arise from a dual transformation of pure gauge theories. Such dual transformations appear also in the strong coupling expansion of lattice gauge theory [1].

It is often stated that LQG is conservative, since it takes classical Einstein gravity as the starting point for quantization, whereas string theory introduces new degrees of freedom and gravity emerges only as an effective theory. While it is true that the definition of LQG is largely inspired by the classical theory, one could also say that it *does* introduce a very new type of degree of freedom: although the quantization starts from fields, it leads to labelled

networks as the fundamental "variables" in the quantum theory. The latter are very different from smooth fields, and have, in fact, some similarities with strings.

The crucial difference between string theory and LQG lies in the role of backgrounds: every string theory comes with a choice of a classical background geometry in which the strings move. To ensure anomaly-freeness and perturbative finiteness, one has to introduce additional structures like supersymmetry and extra dimensions[1]. In contrast, the networks of LQG do not live on any background geometry, but constitute themselves the geometry of space. So far, there is no indication of comparable anomaly conditions[2], and one hopes that the representation of space by labelled networks is enough to render the theory UV finite.

This difference in approach has it that the successes and difficulties of LQG and string theory are in some sense complementary: since string theories are defined on backgrounds, it is relatively simple to derive an effective field theory from them that describes the low-energy regime. The hard task is to show that string theories on different backgrounds belong to a single background-independent theory. LQG, on the other hand, is background-free by construction, but it is far from obvious, if it can lead to effective field theories on classical backgrounds that contain Einstein gravity.

Clearly, the latter is an essential condition if the theory is supposed to have anything to do with reality. More precisely, we can formulate the question as follows: do the transition amplitudes, for a certain class of initial and final states, reduce to the transition amplitudes of an effective field theory on a classical background geometry? The reason why this is so difficult to answer in LQG, lies in what we said about its degrees of freedom: we are not dealing with smooth fields in space anymore, but with networks that *form* space, and it is far from clear how an effective description in terms of fields on a smooth background can be deduced from it.

The aim of this thesis was to develop methods that bring us closer to determining such an effective description of LQG. We call this problem also the problem of the semiclassical limit, because the effective field theory should describe a regime of the theory, where the dominating contributions come from fluctuations near a classical geometry. It encompasses a number of different aspects that can be circumscribed by the following questions:

1. How does the notion of background geometry emerge in loop quantum gravity?
2. In particular, how does the classical notion of time emerge?
3. How can one obtain an effective theory for low-energies from the fundamental definition?
4. Can one do perturbation theory around classical backgrounds?
5. Which quantum states correspond to the vacuum and to gravitons?

These questions already presuppose that the theory is sufficiently well-defined for them to be addressed: for canonical LQG and spin foam gravity this is only partly the case, so we should add as another question:

6. In what way has the theory to be completed, so that we can really begin its physical analysis?

The latter point refers, for example, to the fact that in canonical LQG the definition of dynamics is not clear, and that so far, 4d spin foam models always depended on a choice of triangulation and were not truly background-independent.

[1] There are new proposals for string theories that are free of such consistency requirements [2].

[2] We should add, however, that the formulation of dynamics is not yet sufficiently developed, so that one could really compute anomaly terms.

The research of the Ph.D. project dealt with several of these questions, and can be roughly divided into three phases: the main results are given in the three research articles that constitute chapter 2, 3 and 4 of the thesis [3, 4, 5].

I began with work on the problem of time: there is a new approach, the so-called "general boundary formalism" [6, 7, 8, 9], which extends the usual formulation of dynamics in terms of space-like hypersurfaces to one with closed boundary surfaces. In quantum gravity, this could provide a way to encode "experimental time" in boundary conditions on a sum over geometries [10]. To test the idea, I analyzed if it could be made concrete in the more standard context of scalar field theory: the main result is an evolution equation w.r.t. general boundaries and a proposal for a lattice regularization of the path integral. We report these results in chapter 2. A collaboration with other authors led also to a formal proposal of a vacuum state of LQG, which was published in [10].

The second part of the Ph.D. dealt more directly with the problem of the semiclassical limit; with the question of what corresponds to classical backgrounds and semiclassical fluctuations in LQG. To gain information on this, I took the following strategy: I translated already known states—the vacuum and gravitons of linearized gravity—into states of the canonical loop formalism. The semiclassical properties of the original states turned into properties of the LQG states, and allowed an identification of classical and semiclassical network configurations. The result indicates also a possible way for doing perturbation theory.

In the third part of the Ph.D., I worked on an improvement of the spin foam approach to gravity: the existing models are plagued by the problem that they generally depend on the choice of a triangulation, and that clashes with the requirement of background-independence. I introduced a symmetry principle which requires weights of histories to be independent of the triangulation, and I showed that some models satisfy this condition. In these cases, one is naturally led to a new formulation that is completely background-independent. This can be seen as a contribution to point 6.

The text of the thesis is organized as follows: In sections 1.1 to 1.5, we give a brief introduction to canonical LQG and spin foam gravity. After that, we are in a better position to indicate the contents of the following chapters, which we do in section 1.6. Chapters 2, 3 and 4 contain the three research articles. Since they deal with different problems we have kept them in their original form, with an own introduction and summary at the end. In the final chapter 5, the summary of the entire thesis, we evaluate the results of the Ph.D., and attempt also a critical assessment of the canonical and spin foam approach: we discuss obstacles and new possibilities for the definition and analysis of the theory.

The following introduction only aims at giving basic ideas. For concreteness, we state some technical details, but most of the derivations are omitted, or at best, sketched. We show, for example, the precise form of constraints, but not how they are derived. In some cases this requires very long calculations! We are also sloppy with regard to historical references, since that can be found in more detailed introductions. The reader who is already familiar with LQG may prefer to go directly to section 1.6.

There is a number of good reviews on the subject: the book by Rovelli [11], a more mathematical review by Thiemann [12], the introductions by Perez and Baez [13, 14], and, from a more string theoretical perspective, the review by Nicolai, Peeters and Zamaklar [15]. The connection formulation of classical gravity is reviewed in [16].

1.1 Hamiltonian description of field theory

Let us begin by recalling some general facts about the Hamiltonian description of field theories[3]. The canonical formulation is obtained by Legendre transform from the Lagrangian formalism, and casts the dynamics of the field theory into a space plus time picture. The 4-dimensional spacetime $\mathcal{M}$ is foliated into 3-dimensional hypersurfaces $\Sigma_t \simeq M$ which represent "space" at fixed coordinate time t. The field variables and their derivatives are decomposed according to space and time components. By pulling both the field φ and its time derivative $\dot{\varphi}$ back to M, we can describe a 4d classical solution as an evolution of $(\varphi, \dot{\varphi})$ on M in coordinate time t. The Legendre transform maps $(\varphi, \dot{\varphi})$ into the phase space variables (φ, π), where the momentum density π satisfies

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{\varphi}} \,. \tag{1.1}$$

In general, (1.1) imposes restrictions on the possible values of (φ, π), expressed by the constraints

$$C_k^{(p)}(\varphi, \pi) = 0 \,, \quad k = 1, \ldots, K_p \,. \tag{1.2}$$

These are the so-called primary constraints. Admissible states of the system are restricted to the constraint surface $\Gamma_p \subset \Gamma$ defined by the $C_k^{(p)}$.

The evolution of any phase space function $F(\varphi, \pi)$ is determined by the Poisson bracket

$$\dot{F} = \{F, H_p\} \tag{1.3}$$

with the primary Hamiltonian

$$H_p := H + \int_M N^k C_k^{(p)} \,, \tag{1.4}$$

where H is the standard Hamiltonian $H = \int_M \dot{\varphi}\pi - \mathcal{L}$. The N^k are so far undetermined multiplier fields. Consistency with the Lagrangian formalism requires that states (φ, π) that are initially on Γ_p should remain on it under evolution by H_p. This may imply additional constraints, which in turn have to be checked for consistency, leading possibly to secondary, tertiary etc. constraints. At some point, this iteration procedure (the so-called Dirac-Bergman algorithm) terminates, and one arrives at a finite, total number of constraints

$$C_k = 0 \,, \quad k = 1, \ldots, K \,, \tag{1.5}$$

that are consistent with evolution and define the constraint surface $\Gamma_c \subset \Gamma$.

For understanding the structure of the phase space, it is very useful to distinguish between two classes of constraints: first-class and second-class constraints. First-class constraints are defined by the property that their Poisson brackets with all other constraints vanish on the constraint surface Γ_c. All other constraints, i.e. those that are not first-class, are called second-class.

First-class constraints play a double role in the physics of the phase space: like the other constraints, they specify the constraint submanifold Γ_c, but in addition to that they they also generate flows along Γ_c. This is so, since for any smearing $C_{\rm fc}[N] := \int_M N C_{\rm fc}$, the associated flow vector field $X_{C_{\rm fc}[N]}$ satisfies

$$\mathrm{d}C_k(X_{C_{\rm fc}[N]}) = \{C_k, C_{\rm fc}[N]\} \approx 0 \quad \forall\, k = 1, \ldots, K \,. \tag{1.6}$$

[3] A compact review can be found in [17].

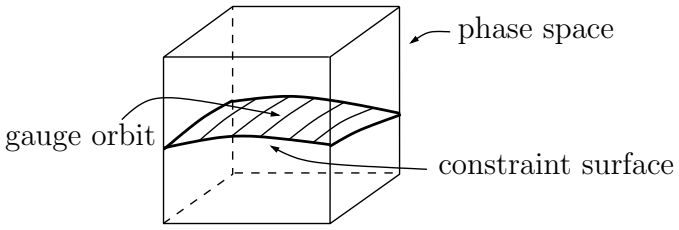


Figure 1.1: Constraint surface of first-class constraint in phase space.

($\approx$ means that the function is weakly zero, i.e. zero on the constraint surface Γ_c.) For the same reason, the multiplier fields N^k of first-class primary constraints drop out and remain undetermined, when imposing the consistency with evolution under H_p. Thus, we can start from a phase space point (φ, π), evolve with two different values of N^k, and both motions represent valid solutions of the classical equations of motion. Often the ambiguity in evolution can be interpreted as a gauge symmetry, and the first-class constraint generates the orbit that relates gauge-equivalent points on the constraint surface. When treating a system with such first-class constraints, we have the option of introducing new constraints that fix the gauge, and reduce the constraint surface Γ_c further such that it intersects each gauge orbit only once.

As opposed to first-class constraints, second-class constraints generate a flow that leads off the constraint surface. In that case, the consistency with evolution fixes the multiplier fields, and no ambiguity appears. In general, one tries to solve for the second-class constraints, which reduces the original phase space Γ to the second-class constraint surface $\Gamma_{\rm sc}$. (This new phase space may still contain constraint surfaces, due to the presence of first-class constraints.)

If $K_{\rm fc}$ and $K_{\rm sc}$ denote the number of first-class and second-class constraints, we see that the number f of physical degrees of freedom per space point is

$$f = \frac{1}{2}\left(\dim\Gamma - 2K_{\rm fc} - K_{\rm sc}\right) = \frac{1}{2}\,\dim\Gamma - K_{\rm fc} - \frac{1}{2}\,K_{\rm sc}\,. \tag{1.7}$$

The admissible motions in phase space are restricted to the total constraint surface Γ_c, which has dimension $\dim\Gamma - K_{\rm fc} - K_{\rm sc}$ for each point in space. By subtracting another time $K_{\rm fc}$, we take account of the gauge orbits, and arrive at the dimension of the physical part of the phase space.

At this point, we should make a clarifying remark on a detail of the Legendre transform that is often skipped over. First-class constraints typically result from a non-dynamical field μ in the Lagrangian that only appears as a multiplying factor of some term C. Therefore, the canonical momentum vanishes:

$$\pi_\mu = \frac{\partial\mathcal{L}}{\partial\dot{\mu}} = 0\,. \tag{1.8}$$

It is standard to say that variation of μ implies the constraint $C = 0$, but this is, strictly speaking, a Lagrange constraint and does not tell us which constraints we receive in the Hamiltonian formalism. If we do the Legendre transform properly, μ will become a canonical variable (and not a multiplier), together with the canonical momentum π_μ, which is also a primary first-class constraint and generates changes in μ.

There are two ways of viewing this result: if we keep (μ, π_μ) as variables of the phase space, any function $\mu(\underline{x}, t)$ is a solution, because its value is arbitrarily shifted by π_μ. In the primary Hamiltonian, μ enters as a coefficient of the term C. Equivalently, we can

exclude (μ, π_μ) from the phase space (by gauge-fixing $\mu = 0$ and phase space reduction), and compensate that by adding a term $\int \mu' C$ to the primary Hamiltonian where now μ' plays the role of a multiplier field.

Dynamics in diff-invariant theories

A second clarifiying remark concerns the relation of dynamics and gauge transformations in diffeomorphism invariant systems. Such systems have the particular property that the Hamiltonian $H = \int_M \dot{\varphi}\pi - \mathcal{L}$ consists only of constraints: the geometry is a variable, and the term in H, which would govern the time evolution for fixed geometry, now takes the form $\int NC$ where C is a first-class constraint, the so-called *Hamiltonian* or *scalar constraint*. N is a multiplier[4]—the lapse field—and determines the proper time interval

$$T_{t_\mathrm{i},t_\mathrm{f}}(\underline{x}) = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \mathrm{d}t\, N(\underline{x}, t) \tag{1.9}$$

that would be measured by a clock at $\underline{x}$ between t_i and t_f. Another first-class constraint that is guaranteed to be present is the *vector constraint* V_a: in the Hamiltonian, it gives rise to the term $\int N_a V^a$, which generalizes the translation generator $\xi_a P^a$ in flat spacetime. Depending on the specific theory, the scalar and vector constraint may be accompanied by additional first-class and second-class constraints.

Due to their first-class property, the scalar and vector constraint can be seen as generators of gauge transformations: the same initial data evolve into different states, depending on the choice of the N and N_a field. In particular, since $\int NC$ generalizes the Hamiltonian of flat spacetime, we can say that the dynamical evolution is a flow along gauge orbits, or in the words of Henneaux & Teitelboim, that the "motion is just the unfolding of a gauge transformation" [18].

It is easy to see how this flow is related to gauge transformations in the covariant (Lagrangian) formalism. Suppose we deal with general relativity, and consider a metric g that is a solution of the Einstein equations. Associated to g we have a conjugate momentum field π. For some time slice Σ_{t_i}, we pull back g and π from $\mathcal{M}$ via Σ_{t_i} to M and obtain the fields $g_\mathrm{i}^{(3)}$ and $\pi_\mathrm{i}^{(3)}$ respectively. Together, they specify a point $p_\mathrm{i} = (g_\mathrm{i}^{(3)}, \pi_\mathrm{i}^{(3)})$ in phase space.

For a given choice of the lapse field $N(\underline{x}, t)$, the smeared constraint $C[N] := \int_M NC$ determines a Hamiltonian vector field $X_{C[N]}$. Exponentiation of $X_{C[N]}$ yields

$$\exp\left(\int_{t_\mathrm{i}}^{t_\mathrm{f}} \mathrm{d}t\; X_{C[N]}\right) = \exp\left(\int_{t_\mathrm{i}}^{t_\mathrm{f}} \mathrm{d}t \int_M N(\underline{x}, t) X_{C(\underline{x})}\right) \tag{1.10}$$

$$= \exp\left(\int_M T_{t_\mathrm{i},t_\mathrm{f}}(\underline{x}) X_{C(\underline{x})}\right). \tag{1.11}$$

This map sends initial states p_i at time t_i into final states $p_\mathrm{f} = (g_\mathrm{f}^{(3)}, \pi_\mathrm{f}^{(3)})$ at time t_f:

$$p_\mathrm{f} = \exp\left(\int_M T_{t_\mathrm{i},t_\mathrm{f}}(\underline{x}) X_C(\underline{x})\right) p_\mathrm{i}\,. \tag{1.12}$$

Since g is a solution, the evolved fields $g_\mathrm{f}^{(3)}$ and $\pi_\mathrm{f}^{(3)}$ correspond to the pull-back of g and π to some hypersurface Σ_{t_f} (see left-hand side of Fig. 1.2). Moreover, we can find a 4d diffeomorphism h that maps Σ_{t_f} into Σ_{t_i} and (g, π) into $(g', \pi') = (h_* g, h_* \pi)$ such that $(g_\mathrm{f}^{(3)}, \pi_\mathrm{f}^{(3)})$ is the pull-back of g' and π' to Σ_{t_i} (see right-hand side of Fig. 1.2). In that sense, the evolution $p_\mathrm{i} \mapsto p_\mathrm{f}$ corresponds to a 4-dimensional gauge-transformation $(g, \pi) \mapsto (g', \pi')$. Equivalently, we could leave the 4d solution g fixed, and apply the inverse diffeomorphism h^{-1} to the hypersurface Σ_{t_i}: then, the final fields $g_\mathrm{f}^{(3)}$ and $\pi_\mathrm{f}^{(3)}$ result from the pull-back of g and π to $h^{-1}(\Sigma_{t_\mathrm{i}})$.

[4] or from the perspective of the Lagrangian formalism, a non-dynamical variable

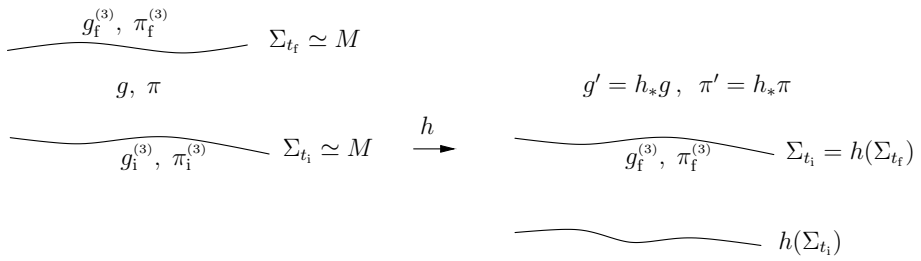


Figure 1.2: Relation between evolution and diffeomorphisms.

We have seen that the scalar constraint can be regarded as a generator of gauge transformations in the canonical formalism, and that it is closely related to gauge transformations in the Lagrangian framework. It should be stressed, however, that among the first-class constraints the scalar constraint plays a very particular role, and that certain caveats apply when we identify the concept of dynamical evolution and gauge transformation.

One should keep in mind that the transformation (1.12) is not a diffeomorphism on $g^{(3)}_{\rm i}$ and $\pi^{(3)}_{\rm i}$ itself: other than a diffeomorphism, the evolution map contains information about the dynamics, which is encoded in the specific form of the action S. Knowledge of the gauge symmetries of S is not sufficient to construct (1.12). Observe also that via equation (1.9) the lapse field N contains diff-invariant information about the 4d solutions that are associated to (1.12): it determines the proper time intervals between points of the initial and final hypersurfaces $\Sigma_{t_{\rm i}}$ and $\Sigma_{t_{\rm f}}$. From that point of view, the gauge ambiguity in the evolution results from the fact that physical time is encoded in $T_{t_{\rm i},t_{\rm f}}(\underline{x})$, and not in the parameter t. We can look at evolution for different proper time intervals, so, in that sense, the freedom in choosing $T_{t_{\rm i},t_{\rm f}}(\underline{x})$ is physical, not gauge.

1.2 Tetrads and SO(1,3) connections

In the connection formulation of gravity, a central role is played by tetrad fields and associated SO(1,3) connections. Below we recap the logic of the tetrad formalism, and how it relates to the more standard description without tetrads.

In differential geometry, we can describe a vector (or tensor) field in terms of coordinate bases $\{\partial/\partial x^a\}$ or non-coordinate bases $\{\theta_I\}$. In the first case, the vector field v corresponds to a section in the tangent bundle of the manifold; in the second case, it is a section in a vector bundle with structure group G, where G is the group that relates local choices of the non-coordinate bases θ_I.

Of course, in both descriptions, the vector field itself is the same, only the component language changes: we have

$$v = v^\alpha\, \partial/\partial x^\alpha = v^I\, \theta_I\,. \tag{1.13}$$

The component vectors are related by

$$v^\alpha = \theta^\alpha{}_I\, v^I\,, \qquad v^I = \theta_\alpha{}^I\, v^\alpha\,, \tag{1.14}$$

where the $\theta_\alpha{}^I$ are the one-forms of the dual basis, i.e.

$$\theta_\alpha{}^I \theta^\alpha{}_J = \delta^I{}_J\,. \tag{1.15}$$

Suppose now that the manifold is 4-dimensional and carries a metric g. When the θ_I are required to be orthonormal, they are called tetrads e_I (or triads in 3 dimensions). The

one-forms $e_\alpha{}^I$ are referred to as co-tetrads. For Lorentzian g, we have

$$e^\alpha{}_I e^\beta{}_J\, g_{\alpha\beta} = \eta_{IJ}\,, \qquad I, J = 0, 1, 2, 3, \tag{1.16}$$

and the component fields v^I are (local) sections in a $SO(1,3)$ vector bundle.

Connections on principal fibre bundles determine covariant derivatives on the associated vector bundles. For the tangent bundle, whose structure group is GL(4,$\mathbb{R}$), a covariant derivative has the form

$$D^{\rm GL}_\alpha v^\beta = \partial v^\beta + C^\beta{}_{\alpha\gamma}\, v^\gamma\,. \tag{1.17}$$

The $C^\beta{}_{\alpha\gamma}$'s arise from a local gl(4,$\mathbb{R}$)-valued connection form

$$C = C_\alpha\, {\rm d}x^\alpha\,. \tag{1.18}$$

The one-form component C_α is an element of gl(4,$\mathbb{R}$), and the $C^\beta{}_{\alpha\gamma}$ denote its matrix entries. $D^{\rm GL}_\alpha$ is assumed to be torsion-free, so the lower indices of $C^\beta{}_{\alpha\gamma}$ are symmetric. For the SO(1,3) vector bundle, we get a covariant derivative

$$D^{\rm SO}_\alpha v^I = \partial v^I - A_\alpha{}^I{}_J\, v^J\,, \tag{1.19}$$

with the so(1,3)-valued connection form

$$A = A_\alpha\, {\rm d}x^\alpha\,. \tag{1.20}$$

The $A_\alpha{}^I{}_J$ are the matrix elements of $A_\alpha \in {\rm so}(1,3)$.

Here, we want the two covariant derivatives to be equivalent, that is

$$D^{\rm GL}_\alpha v^\beta \stackrel{!}{=} e^\beta{}_I\, D^{\rm SO}_\alpha v^I\,. \tag{1.21}$$

If we demand furthermore that $D^{\rm GL}_\alpha$ be compatible with the metric g, i.e.

$$D^{\rm GL} g_{\alpha\beta} = 0\,, \tag{1.22}$$

the GL(4,$\mathbb{R}$) connection is fixed to be the Levi-Civita connection: i.e. $C^\beta_{\alpha\gamma} = \Gamma^\beta_{\alpha\gamma}$ and $D^{\rm GL}_\alpha = \nabla_\alpha$.

Together with (1.21), the metric condition fixes also the SO(1,3) connection: according to (1.21), we have

$$e^\beta{}_J D^{\rm SO}_\alpha \delta^J{}_I = D^{\rm GL}_\alpha e^\beta{}_I + e^\beta{}_J\, A_{\alpha I}{}^K\, \delta^J{}_K \tag{1.23}$$

but

$$D^{\rm SO}_\alpha \delta^J{}_I = e_\alpha{}^J\, e^\beta{}_I\, \nabla_\alpha \delta^\alpha{}_\beta = 0\,. \tag{1.24}$$

Therefore, A satisfies

$$A_{\alpha I}{}^J = -e_\beta{}^J\, \nabla_\alpha e^\beta{}_I =: \omega_{\alpha I}{}^J\,, \tag{1.25}$$

which is the defining equation for the connection form ω—the so-called spin-connection; its components are sometimes referred to as Ricci rotation coefficients.

The curvature 2-forms of C and A are given by

$$\mathcal{D}_{[\alpha}\mathcal{D}_{\beta]} v_\gamma = F_{\alpha\beta\gamma}{}^\delta\, v_\delta\,, \qquad \mathcal{D}_{[\alpha}\mathcal{D}_{\beta]} v_I = F_{\alpha\beta I}{}^J\, v_J\,. \tag{1.26}$$

where

$$F_{\alpha\beta}{}^\gamma{}_\delta = \partial_{[\alpha}\Gamma^\gamma{}_{\beta]\delta} + [\Gamma_\alpha, \Gamma_\beta]^\gamma{}_\delta\,, \qquad F_{\alpha\beta I}{}^J = \partial_{[\alpha}\omega^I{}_{\beta]J} + [\omega_\alpha, \omega_\beta]^I{}_J\,. \tag{1.27}$$

They are related by

$$F_{\alpha\beta}{}^{\gamma\delta} = e^\gamma{}_I\, e^\delta{}_J\, F_{\alpha\beta}{}^{IJ}\,, \tag{1.28}$$

and equal the Riemann curvature R when $C = \Gamma$ (or $A = \omega$).

In the language of bundles, tensors with mixed $I, J \ldots$ and $\alpha, \beta \ldots$ indices correspond to sections in tensor products of the GL(4,$\mathbb{R}$) and SO(1,3) vector bundles. The tetrads serve to transform SO(1,3) indices into GL(4,$\mathbb{R}$) indices, and vice versa. On such tensor products, we have the combined covariant derivative

$$\mathcal{D}_\alpha = \partial v^\beta + C^\beta{}_{\alpha\gamma}\, v^\gamma - A_\alpha{}^I{}_J\, v^J \,. \tag{1.29}$$

We have seen that compatibility between the connections (eqn. (1.21)) and with the metric (eqn. (1.22)) implies equations (1.23) and (1.24). The latter give

$$\mathcal{D}_\alpha e^\beta{}_I = 0 \tag{1.30}$$

for the total covariant derivative. Conversely, (1.30) leads immediately to (1.21) and (1.22). Therefore, the single equation (1.30) encodes both the metric property of the GL(4,$\mathbb{R}$) connection ($C = \Gamma$), and its equivalence to the SO(1,3) connection A.

1.3 Hamiltonian formulations of gravity

In the following subsections, we present several Hamiltonian descriptions of classical gravity that lead up to the Ashtekar-Barbero formalism. These formulations differ in the actions and variables that are used as a starting point for the Legendre transform. Their solutions are equivalent, however.

It is always assumed that the 3d slices $\Sigma_t \simeq M$ are compact without boundary, so no boundary terms appear. We reserve the letter C for the scalar constraint. We use Greek indices at the beginning of the alphabet (i.e. $\alpha, \beta, \gamma \ldots$) as spacetime indices and Latin indices $a, b, c \ldots$ for space. For indexing tetrads, we employ capital letters in the middle of the alphabet ($I, J, K \ldots$), and $i, j, k, \ldots$ for triads. η_{IJ} stands for the Minkowski metric w.r.t. tetrad indices.

$\mathcal{D}$ denotes torsion-free covariant derivatives of the form (1.29), which act on both spacetime and tetrad indices. ∇_a stands for the covariant derivative w.r.t. the spatial part of the Levi-Civita connection Γ. Sometimes there appear covariant derivatives that refer only to SO(1,3), SO(4) or a subgroup of it: in that case, we write D_α for the derivative. Curvatures are denoted by F (for general connections), and by R for the Levi-Civita and spin connection Γ and ω. The anti-symmetrizing bracket $[\ldots]$ is weighted with 1, i.e. $T_{[\alpha\beta]} = T_{\alpha\beta} - T_{\beta\alpha}$. We set $\kappa = 8\pi G/c^2$, where G is Newton's constant.

1.3.1 Einstein-Hilbert action

. . . with metrics

The standard metric formulation of gravity can be derived from the Einstein-Hilbert Lagrangian

$$\mathcal{L}_{\mathrm{EH}} = \sqrt{|g|}\,(R + 2\lambda) \,. \tag{1.31}$$

Here, g denotes the determinant $\det(g_{\alpha\beta})$ and λ is the cosmological constant. Variation of $g_{\alpha\beta}$ yields the Einstein equation

$$R_{\alpha\beta} - \frac{1}{2}\, g_{\alpha\beta}\, R - \lambda\, g_{\alpha\beta} = 0 \,. \tag{1.32}$$

The Legendre transform leads to the so-called ADM formulation of gravity: the canonical variables are g_{ab} and $\pi^{ab}(\underline{x})$ where g_{ab} is the 3-metric, and $\pi^{ab}(\underline{x})$ is related to the extrinsic curvature by

$$\pi^{ab}(\underline{x}) = \sqrt{g}\,\left(K^{ab} - K\, g_{ab}\right) \,. \tag{1.33}$$

The Poisson brackets read

$$\left\{g_{ab}(\underline{x}), \pi^{cd}(\underline{y})\right\} = \delta_a{}^c \delta_b{}^d \, \delta(\underline{x} - \underline{y}) \,. \tag{1.34}$$

The Hamiltonian contains the two first-class constraints

$$V^a = \nabla_b \left(\sqrt{g}^{-1} \pi^{ab}\right) , \tag{1.35}$$

$$C = \sqrt{g}^{-1} \left(\pi^{ab}\pi_{ab} - \frac{1}{2}\,\pi^2\right) - \sqrt{g}\left(R + 2\lambda\right) , \tag{1.36}$$

the so-called vector constraint V^a, which generates spatial diffeomorphisms in M, and the scalar constraint C that determines the dynamics. The R and g on the right-hand side refer to the 3-dimensional curvature and metric. When counting degrees of freedom according to (1.7), we get 6 for the phase space, and -4 for the first-class constraints, so we have 2 degrees of freedom per space point.

. . . with tetrads

The information about the metric $g_{\alpha\beta}$ can be equivalently encoded in tetrads and cotetrads, since

$$g_{\alpha\beta} = e_\alpha{}^I e_\beta{}^J \, \eta_{IJ} \,. \tag{1.37}$$

The only difference is the additional SO(1,3) gauge symmetry: we can locally Lorentz boost and rotate the bases e_I without changing the geometry. When expressed in terms of tetrads, the Lagrangian density (1.31) takes the form

$$\mathcal{L}_{\text{EH}} = e \left(e^\alpha{}_I e^\beta{}_J \, R_{\alpha\beta}{}^{IJ} + 2\lambda\right) , \tag{1.38}$$

where $R_{\alpha\beta}{}^{IJ}$ is the curvature of the spin-connection ω. We lower and raise tetrad indices with the Minkowski metric η_{IJ}, so $e^{\alpha I} := \eta^{IJ} \, e^\alpha{}_J$ and $e_{\alpha I} := \eta_{IJ} \, e_\alpha{}^J$. When writing e, we mean the determinant $\det(e^\alpha{}_I)$, which is the same as $\sqrt{g}$.

The equations of motion are

$$e^\alpha{}_K e^\gamma_I e^\delta{}_L R_{\gamma\delta}{}^{KL} - \left(\lambda + \frac{1}{2} e^\gamma{}_K e^\delta{}_L R_{\gamma\delta}{}^{KL}\right) e^\alpha{}_I \,. \tag{1.39}$$

After the Legendre transform[5] the variables are the triad e_{ai}, $a, i = 1, 2, 3$, and the canonical momentum

$$\pi^a{}_i = \frac{1}{2}\, e \, e^a{}_j \left(K_{ij} - \delta_{ij} K\right) , \tag{1.40}$$

with Poisson brackets

$$\left\{e_{ai}(\underline{x}), \pi^{bj}(\underline{y})\right\} = \delta_a{}^b \delta_i{}^j \, \delta(\underline{x} - \underline{y}) \,. \tag{1.41}$$

K_{ij} is the extrinsic curvature in the triad basis.

The constraint analysis gives an additional constraint G_{ij} that takes account of the gauge freedom in the triads:

$$G_{ij} = e_{a[i} \, \pi^a{}_{j]} \tag{1.42}$$

$$V_a = e_{ai} \nabla_b \pi^{bi} \tag{1.43}$$

$$C = e^{-1} \left(\pi^{ij}\pi_{ij} - \frac{1}{2}\left(\pi^i{}_i\right)^2\right) - \frac{1}{4}\, eR - \frac{1}{2}\, e\lambda \,. \tag{1.44}$$

[5]The calculation can be found in [19].

Here,

$$\pi^{ij} := e_a{}^i \pi^{aj} \,. \tag{1.45}$$

Note that the original SO(1,3) gauge group has been partially gauge-fixed to SO(3). Let us again count the degrees of freedom: 9 for the phase space, 7 from first-class constraints (4 corresponding to diffeomorphisms, 3 corresponding to SO(3) gauge transformations), which leaves 2 physical degrees of freedom.

With hindsight to later, it is useful to reformulate (1.41)-(1.44) in terms of the *densitized* triad

$$E^a{}_i = e\, e^a{}_i \,. \tag{1.46}$$

Its canonical momentum is

$$K_b{}^j := K_{bc}\, e^{ci} \,, \tag{1.47}$$

The Poisson brackets and constraints become

$$\left\{ E^a_i(\underline{x}), K^k_b(\underline{y}) \right\} = \frac{\kappa}{2}\, \delta^a_b \delta^k_i\, \delta(\underline{x} - \underline{y}) \,, \tag{1.48}$$

$$\begin{aligned} G_{ij} &= K_{a[i}\, E^a{}_{j]} \,, & (1.49)\\ V_a &= \nabla_{[b} \left(K_{a]}{}^i\, E^b{}_i \right) \,, & (1.50)\\ C &= -\frac{1}{\sqrt{|E|}} K_a{}^{[i} K_b{}^{j]} E^a{}_i E^b{}_j - \sqrt{|E|}\, (R + 2\lambda) \,. & (1.51) \end{aligned}$$

E denotes the determinant $\det(E^a{}_i)$. We call the pair (E, K) the *extended* ADM variables.

1.3.2 Hilbert-Palatini action

The general strategy of LQG is to put classical gravity into a form where the connection is a configuration variable, and to start quantization from there. So far, the curvature in (1.38) is the curvature of the spin-connection, which is not a variable, but fixed by the value of the tetrads. The key input that allows for a connection formulation is the following: one can turn the spin-connection into an independent variable and still maintain equivalence with GR!

When we replace the spin-connection ω with an arbitrary SO(1,3) connection A, we obtain the so-called Hilbert-Palatini Lagrangian:

$$\mathcal{L}_{\mathrm{HP}} = e \left(e^\alpha{}_I e^\beta{}_J F_{\alpha\beta}{}^{IJ} + 2\lambda \right) \,. \tag{1.52}$$

The equivalence to the Einstein-Hilbert action can be seen as follows: the variation of e proceeds as before, and gives

$$e^\alpha{}_K e^\gamma_I e^\delta{}_L F_{\gamma\delta}{}^{KL} - \left(\lambda + \frac{1}{2} e^\gamma{}_K e^\delta{}_L F_{\gamma\delta}{}^{KL} \right) e^\alpha{}_I = 0 \tag{1.53}$$

The only difference is that R is replaced by F. For the variation in ω, we use

$$\delta R_{\alpha\beta}{}^{IJ} = D_{[\alpha} \delta\omega_{\beta]}{}^{IJ} \,, \tag{1.54}$$

and partial integration, which yields

$$\mathcal{D}_\alpha \left(e\, e^{[\alpha}{}_I e^{\beta]}{}_J \right) = 0 \,. \tag{1.55}$$

It can be shown that the latter is equivalent to

$$\mathcal{D}_{[\alpha} e_{\beta]}{}^I = 0 \,. \tag{1.56}$$

This is enough to solve for the connection A, and implies $A = \omega$. Thus, we can put $F = R$ in (1.53), and get back to the Einstein equation (1.39).

Unfortunately, (1.52) does not lead to a canonical formulation in terms of the connection. Its constraint analysis is complicated and involves second-class constraints. For 18 ${A_a}^{IJ}$'s, we have only 10 first-class constraints (4 corresponding to diffeomorphisms, 6 corresponding to Lorentz gauge transformations), so there have to be 12 second-class constraints to reduce the physical degrees of freedom to 2. When solving these second-class constraints, one is forced back to the ADM formalism with densitized triads (see (1.46) to (1.51)).

1.3.3 Self-dual action

The appearance of second-class constraints can be seen as the consequence of a surplus of 6 ${A_a}^{IJ}$'s in the Hilbert-Palatini action. Remarkably, their number can be reduced by a trick—another modification of the action, which leaves only the self-dual part of the connection. With these self-dual variables, the balance with the first-class constraints is recovered and no second-class constraints arise. The Legendre transform leads to a canonical formulation with a connection—the so-called Ashtekar connection.

For simplicity, we describe the idea for Riemannian gravity, and state in the end why it becomes more problematic for the Lorentzian case. The modified Lagrangian reads

$$\mathcal{L}_{\text{sd}} = e\, {e^\alpha}_I {e^\beta}_J \left({F_{\alpha\beta}}^{IJ} + \star {F_{\alpha\beta}}^{IJ} \right) + 2e\,\lambda \tag{1.57}$$

Here, $\star$ stands for the Hodge star operation on SO(4) indices, i.e.

$$\star\, {F_{\alpha\beta}}^{IJ} = \frac{1}{2} {\epsilon^{IJ}}_{KL}\, {F_{\alpha\beta}}^{KL}\,. \tag{1.58}$$

We have

$$\star^2 = \mathbb{1}\,, \tag{1.59}$$

since the signature s is positive. Lie algebra elements M that satisfy

$$\star\, M = \pm\, M\,, \tag{1.60}$$

are called self-dual and anti-self-dual respectively. It is well-known from quantum field theory that the Lie algebra so(1,3) splits into two ideals. So does also so(4), namely,

$$\text{so}(4) = \text{so}^{(+)}(4) \oplus \text{so}^{(-)}(4)\,, \tag{1.61}$$

where the two parts are isomorphic to su(2) and given by the self-dual and anti-self-dual subspaces. The projector onto the self-dual part is

$$P^{(\pm)} = \frac{1}{2}\left(\mathbb{1} \pm \star\right)\,, \tag{1.62}$$

and we write $M^{(\pm)} = P^{(\pm)}\, M$ for any Lie algebra element M. It is immediate from (1.27) that

$$F^{(+)}(A) = F(A^{(+)})\,. \tag{1.63}$$

We see that the additional term in (1.57) has just the effect of projecting out the anti-self-dual part of A:

$$\mathcal{L}_{\text{sd}} = 2e \left({e^\alpha}_I {e^\beta}_J {F^{(+)IJ}_{\alpha\beta}} + \lambda \right) \tag{1.64}$$

The 6 equations of self-duality reduce the number of ${A_a}^{IJ}$'s from 18 to 12, which is exactly needed to balance the first-class constraints.

Let us check the equivalence with GR: variation of A gives

$$\mathcal{D}^{(+)}_{[\alpha} e_{\beta]}{}^I = 0\,, \tag{1.65}$$

and implies that $A^{(+)}$ equals the self-dual part $\omega^{(+)}$ of the spin-connection. By plugging this back into the Lagrangian, we obtain

$$\mathcal{L}_{\rm sd}{}_{\big|A^{(+)}=\omega^{(+)}} = e\left(e^\alpha{}_I e^\beta{}_J R_{\alpha\beta}{}^{IJ} - e^\alpha{}_I e^\beta{}_J \star R_{\alpha\beta}{}^{IJ} + 2\lambda\right)\,. \tag{1.66}$$

Next we use the identity

$$e\, e^{\alpha[I}\, e^{\beta J]} = \frac{1}{2}\epsilon^{\alpha\beta\gamma\delta}\epsilon^{IJMN} e_{\gamma M} e_{\delta N} \tag{1.67}$$

to show that

$$\begin{aligned} e\, e^\alpha{}_I\, e^\beta{}_J\, \epsilon^{IJKL}\, R_{\alpha\beta KL} &= \frac{1}{4}\epsilon^{\alpha\beta\gamma\delta}\epsilon_{IJ}{}^{MN}\epsilon^{IJ}{}_{KL}\, e_{\gamma M} e_{\delta N} R_{\alpha\beta}{}^{KL} && (1.68)\\ &= \epsilon^{\alpha\beta\gamma\delta}\epsilon^{IJKL} R_{\alpha\beta\gamma\delta} && (1.69)\\ &= 0\,. && (1.70)\end{aligned}$$

The last line follows from the properties of the Riemann tensor. Therefore, the second term in (1.66) vanishes, and we recover the Einstein-Hilbert Lagrangian.

The Legendre transform is straightforward and leads to the Poisson brackets

$$\left\{A_a{}^{IJ}(\underline{x}), E^b{}_{KL}(\underline{y})\right\} = \kappa\, P^{(+)IJ}{}_{KL}\, \delta_a{}^b \delta(\underline{x}-\underline{y}) \tag{1.71}$$

Here and below we drop the superscript $^{(+)}$ on the self-dual connection. The canonical momentum $E^a{}_{IJ}$ is related to the 3-metric via

$$E^a{}_{IJ} E^{bIJ} = g\, g^{ab}\,. \tag{1.72}$$

It is convenient to use a basis of self-dual matrices $T^{(+)}_i$, $i = 1, 2, 3$, (where $T^{(+)IJ}_i T^{(+)}_j{}_{IJ} = \delta_{ij}$), and write

$$A_a{}^{IJ} = A_a{}^i T^{(+)IJ}_i\,, \qquad E^b{}_{KL} = E^b{}_j T_{KL}{}^j\,. \tag{1.73}$$

In terms of the components $A_a{}^i$ and $E^b{}_j$, the Poisson brackets become

$$\left\{A_a{}^i, E^b{}_j\right\} = \kappa\, \delta_a{}^b \delta^i{}_j \delta(\underline{x}-\underline{y})\,. \tag{1.74}$$

The constraints have the simple form

$$\begin{aligned} G_i &= D_a E^a{}_i && (1.75)\\ V_a &= E^b{}_i F_{ab}{}^i && (1.76)\\ C &= \epsilon_{ijk} E^{ai} E^{bj} F_{ab}{}^k + 2g\lambda\,. && (1.77)\end{aligned}$$

D and $F = F^i\, T^{(+)}_i$ denote the covariant derivative and curvature w.r.t. A.

When SO(4) is replaced by SO(1,3), equation (1.59) changes to

$$\star^2 = -\mathbb{1}\,. \tag{1.78}$$

and the eigenvalues of $\star$ become $\pm\,\mathrm{i}$: the associated eigenvectors are again referred to as self-dual and anti-self-dual. The "self-dual" Lagrangian is now

$$\mathcal{L}_{\rm sd} = e\left(e^\alpha{}_I e^\beta{}_J F_{\alpha\beta}{}^{IJ} - \mathrm{i}\star F_{\alpha\beta}{}^{IJ}\right) + 2e\,\lambda\,. \tag{1.79}$$

Equivalence to GR follows in the same way as for SO(4). One can even allow complex values for the connection, since (1.65) determines A as a function of the real tetrads. Thus, we obtain a covariant formulation of GR with complex so(1,3,$\mathbb{C}$) connections and real tetrads. The canonical formulation leads to *complex* phase space variables ${A_a}^i$ and ${E^b}_j$ where A is a $\mathrm{so}^{(+)}(1,3,\mathbb{C})$ connection. The constraints have the same form as for SO(4), and additional reality conditions ensure that solutions correspond to real metrics of general relativity[6].

The disadvantage of this approach is that the connection takes values in a Lie algebra of a non-compact gauge group, and that one has to drag along the reality conditions in order to recover real GR. A mathematically rigorous quantization for non-compact gauge groups poses severe technical difficulties, and the quantum implementation of the reality conditions is complicated. Neither of these problems has been resolved so far. As a result, the main part of present-day LQG is based on a different canonical formulation: it has the special property of giving an SU(2) connection in the Lorentzian case! To arrive at it, we need a last modification of the gravitational action: the introduction of the Immirzi parameter.

1.3.4 Action with Immirzi parameter

In the self-dual Lagrangian (1.79), we replace the projector

$$P^{(+)} = \frac{1}{2}\left(\mathbb{1} \,-\, \mathrm{i}\star\right) \tag{1.80}$$

by the map

$$P^{(\beta)} = \frac{1}{2}\left(\mathbb{1} \,-\, \frac{1}{\beta}\star\right) . \tag{1.81}$$

where β is an arbitrary non-zero and *real* parameter. This yields

$$\begin{aligned} \mathcal{L} &= e\, e^\alpha{}_I e^\beta{}_J \left(F_{\alpha\beta}{}^{IJ} - \frac{1}{\beta}\star F_{\alpha\beta}{}^{IJ} + 2\lambda\right) && (1.82)\\ &= e\left(e^\alpha{}_I e^\beta{}_J\, P^{(\beta)}\, F_{\alpha\beta}{}^{IJ} + 2\lambda\right) . && (1.83) \end{aligned}$$

For $\beta = -\mathrm{i}$, the map $P^{(\beta)}$ would project out the self-dual part and we would return to the self-dual formulation. Here, we restrict β to be real, and $P^{(\beta)}$ becomes invertible. Thus, variation of A implies

$$\mathcal{D}_{[\alpha} e_{\beta]}{}^I = 0 . \tag{1.84}$$

By the same arguments as before, we conclude that (1.83) is equivalent to the Einstein-Hilbert Lagrangian.

The transition to the Hamiltonian formulation is complicated, since it involves second-class constraints. This time, however, the solving of these constraints does not lead back to the ADM formalism. The space-time part of $P^\beta A_a$ becomes the new canonical variable:

$$A'_{ai} := P^\beta A_{a0i} = A_{a0i} - \frac{1}{2\beta}\,\epsilon_{0ijk}\, A_a{}^{jk} . \tag{1.85}$$

Its conjugate momentum is the densitized triad

$${E^i}_k = e\, e^i{}_k . \tag{1.86}$$

The spatial part of A_a takes the value of the spin-connection, i.e.

$$A_a{}^{ij} = \omega_a{}^{ij} . \tag{1.87}$$

[6]This is the original Asthekar formulation of gravity that provides the historical starting point for the development of loop quantum gravity [20, 21, 22].

The gauge group is reduced from SO(1,3) to SO(3) by imposing the so-called time gauge: this means that at every point, the vector e_0 is normal to the foliation slice Σ_t, and that the remaining e_i span the tangent space of Σ_t.

For any solution, the entire A equals the spin-connection, so

$$A_{ai0} = \omega_{ai0} = e^{\beta}{}_i \nabla_a e_{\beta 0} \,. \tag{1.88}$$

The latter is the extrinsic curvature K_{ai} of the Σ_t-slices. Together with (1.85), this implies

$$A'_{ai} = K_{ai} - \frac{1}{2\beta}\, \epsilon_{ijk}\, \omega_a{}^{jk} \,. \tag{1.89}$$

In the so(3) basis $(J_i)^j{}_k = \epsilon_i{}^j{}_k$, the connection ω has the components

$$\omega_a{}^i = \frac{1}{2}\, \epsilon^i{}_{jk}\, \omega_a{}^{jk} \,, \tag{1.90}$$

Under SO(3) gauge transformations, ω^i transforms like a component of a connection, and K_i transforms like a vector. Therefore,

$$A'_{ai} = K_{ai} - \frac{1}{\beta}\, \omega_{ai} \,, \tag{1.91}$$

transforms likewise as a so(3) connection. Due to the isomorphism between so(3) and su(2), one may interpret it also as the components of an su(2) connection

$$A'_a = A'^{\,i}_a \, \mathrm{i}\sigma_i/2 \,. \tag{1.92}$$

(The σ_i, $i = 1, 2, 3$, stand for the Pauli matrices.) The Poisson brackets read

$$\left\{ A'^{\,i}_a(\underline{x}), E^b{}_j(\underline{y}) \right\} = -\kappa\, \delta_a{}^b \delta^i{}_j \delta(\underline{x} - \underline{y}) \,. \tag{1.93}$$

It is convenient to introduce the rescaled connection

$$A_a{}^i := \beta\, A'^{\,i}_a \,, \tag{1.94}$$

for which

$$A_{ai} = \beta\, K_{ai} - \omega_{ai} \,, \tag{1.95}$$

and

$$\left\{ A_a{}^i(\underline{x}), E^b{}_j(\underline{y}) \right\} = -\kappa\beta\, \delta_a{}^b \delta^i{}_j \delta(\underline{x} - \underline{y}) \,. \tag{1.96}$$

In terms of A and E, the constraints read

$$\begin{aligned} G_i &= D_a E^a{}_i := \partial_a E^a{}_i + \epsilon_i{}^{jk} A_{aj} E^a{}_k \,, & (1.97)\\ V_a &= E^b{}_i F_{ab}{}^i \,, & (1.98)\\ C &= \epsilon^{ij}{}_k\, E^a{}_i E^b{}_j \left[F_{ab}{}^k - \left(\beta^2 + 1\right) R_{ab}{}^k \right] \,. & (1.99) \end{aligned}$$

We see from (1.96) and (1.48) in sec. 1.31 that (1.95) represents a canonical transformation from the extended ADM variables $(E^a{}_i, K_b{}^j)$ to the new variables $(A_a{}^i, E^b{}_j)$. We refer to them as the Ashtekar-Barbero variables. They were first introduced by Barbero, using the canonical transformation from $(E^a{}_i, K_b{}^j)$ [23], and later Holst derived them from the Lagrangian (1.83) [24, 25].

The merit of this formulation is that it describes Lorentzian gravity in terms of a connection with compact gauge group. As a result, its quantization made much more progress than in the self-dual, non-compact case. The compactness comes at the price of a scalar constraint that is considerably more complicated than for self-dual gravity. Moreover, the connection (1.95) does not have a simple geometric interpretation.

1.4 Canonical loop quantum gravity

1.4.1 Dirac quantization

The quantization scheme of canonical LQG has two main characteristics: it follows the program of Dirac quantization, and it treats holonomies and electric fluxes as the fundamental quantities of the theory.

The Dirac program prescribes rules for quantizing canonical systems with first-class constraints: to start with, one introduces a kinematic Hilbert space $\mathcal{H}_{\mathrm{kin}}$ that corresponds to a quantization of all phase space variables. The first-class constraints C_i are interpreted as generators of gauge transformations: they are translated into operators $\hat{C}_i$ that generate gauge transformations on states in $\mathcal{H}_{\mathrm{kin}}$. The invariance condition $\hat{C}_i\Psi = 0$ determines the subspace $\mathcal{H}_{\mathrm{phys}}$ of *physical* states.

In the case of LQG, the constraints are imposed in the following order:

$$\mathcal{H}_{\mathrm{kin}} \xrightarrow{G_i} \mathcal{H}_0 \xrightarrow{V_a} \mathcal{H}_{\mathrm{diff}} \overset{C}{\dashrightarrow} \mathcal{H}_{\mathrm{phys}} \tag{1.100}$$

The imposition of SU(2) gauge invariance leads to the gauge-invariant Hilbert space $\mathcal{H}_0$. Then, the invariance under spatial diffeommorphism is implemented, giving the diff-invariant Hilbert space $\mathcal{H}_{\mathrm{diff}}$. The final step consists in finding solutions (or at least approximations) to the equation $\hat{C}\Psi = 0$.

In the classical theory, the first-class constraints play a double role, as they define both the constraint surface and the gauge orbits on that surface. A priori, it is not obvious why in the quantum theory this should translate into a single condition, namely, that physical states be annihilated by the constraint operators. An intuitive explanation for this can be found in the path integral formulation: consider, for example, a formal path integral

$$\int DN \int D\underline{N} \int D\Lambda \int DA \int DE \tag{1.101}$$

$$\times\ \Psi_1^*[A|_{t=1}] \exp\left\{\frac{\mathrm{i}}{\hbar}\int_0^1 \mathrm{d}t \int_M \mathrm{d}^3x\ \left(\dot{A}E - NC - N^a V_a - \Lambda_i G^i\right)\right\} \Psi_0[A|_{t=0}] \tag{1.102}$$

over the Ashtekar-Barbero variables (A, E) and multipliers N, $\underline{N}$ and Λ. The boundary values at $t = 0$ and $t = 1$ are weighted by functionals Ψ_0 and Ψ_1. After the transition to the operator formalism, the same amplitude takes the form

$$\int DN \int D\underline{N} \int D\Lambda\ \langle\Psi_1| \exp\left\{\frac{\mathrm{i}}{\hbar}\int_{[0,1]\times M} \mathrm{d}^4x\ \left(N\hat{C} - N^a \hat{V}_a - \Lambda_i \hat{G}^i\right)\right\} |\Psi_0\rangle\,. \tag{1.103}$$

The integration over the multipliers projects out the part of the states that is annihilated by the constraint operators. What is projected out, cannot influence transition amplitudes, is therefore not measurable, and must correspond to unphysical degrees of freedom. Thus, it seems reasonable to impose the condition

$$\hat{C}\Psi = \hat{V}_a\Psi = \hat{G}^i\Psi = 0 \tag{1.104}$$

on physical states.

1.4.2 Holonomies and electric fluxes

The other central aspect of LQG is the use of connection variables, and the choice of holonomies and fluxes as the basic objects of the theory. In that sense, the loop approach is very similar to (Hamiltonian) lattice gauge theory. In the context of Ashtekar gravity, the holonomy variables were introduced by Jacobson, Rovelli and Smolin [26, 27].

Consider a smooth SU(2) connection A on the space manifold M, and let e be a smooth oriented path in M. In the following, we refer to such paths as *edges*. For the given connection, the edge e defines a holonomy

$$\overline{A}_e := \mathcal{P} \exp\left(\mathrm{i} \int_e \mathrm{d}s \; \dot{e}^a(s) A_a^i(\underline{e}(s)) J_i \right) \in SU(2) \,. \tag{1.105}$$

Here, the J_i, $i = 1, 2, 3$, denote a basis of su(2), and $\mathcal{P}$ indicates path-ordering. If we know the holonomies for all edges e in the manifold, we have complete knowledge of the field A. Thus, we can describe A equivalently by the map

$$\overline{A} : E \to \mathrm{SU}(2) \,, \quad e \mapsto \overline{A}_e \,, \tag{1.106}$$

that associates to each edge e the holonomy $\overline{A}_e \in \mathrm{SU}(2)$. Clearly, this map satisfies the conditions

$$\overline{A}_{e_2} \overline{A}_{e_1} = \overline{A}_{e_2 \circ e_1} \qquad \text{and} \qquad \overline{A}_{e^{-1}} = \overline{A}_e^{-1} \,, \tag{1.107}$$

where e^{-1} and $e_2 \circ e_1$ denote the inverse and composition of edges.

The definition (1.105) refers to the fundamental representation, but we can extend it easily to any other representation of SU(2). Let R_j be the spin j representation, and choose a basis ${}^jJ_i$, $i = 1, 2, 3$, for the associated representation of the Lie algebra. Then, we define the holonomy of A in the j-representation by

$$U_e^j[A] := \mathcal{P} \exp\left(\mathrm{i} \int_e \mathrm{d}s \; \dot{e}^a(s) A_a^i(\underline{e}(s)) \, {}^jJ_i \right) \,. \tag{1.108}$$

Note that

$$U_e^j[A] = R_j(\overline{A}_e) \qquad \text{and} \qquad U_e^{1/2}[A] = \overline{A}_e \,. \tag{1.109}$$

Let us next explain what we mean by an electric flux. In the classical theory, the conjugate momenta of A are smooth triad fields

$$E_i(\underline{x}) = E_i^a(\underline{x}) \frac{\partial}{\partial x^a} \,, \quad i = 1, 2, 3 \,. \tag{1.110}$$

If we dualize a given E_i-field with the tensor density ϵ_{abc}, we obtain a 2-form $\epsilon_{abc} \, E^{ci}$. The latter has density weight 0, since the weight $+1$ of E cancels the weight -1 of ϵ. Thus, we can pull-back $\epsilon_{abc} \, E^{ci}$ to a surface $\Sigma \subset M$, and integrate it over it. The resulting quantity is

$$E_i(\Sigma) := \int_\Sigma f^* \epsilon_{abc} \, E^c{}_i \,. \tag{1.111}$$

where f^* denotes the pull-back of the embedding $f : \Sigma \to M$. We call $E_i(\Sigma)$ the electric flux of E_i through the surface Σ. In coordinates, it takes the form

$$E_i(\Sigma) = \int_\Sigma \mathrm{d}^2\sigma \; \frac{\partial f^a(\underline{\sigma})}{\partial \sigma^1} \frac{\partial f^b(\underline{\sigma})}{\partial \sigma^2} \, \epsilon_{abc} \, E^c{}_i(f(\sigma)) = \int_\Sigma \mathrm{d}^2\sigma \; n_c(\underline{\sigma}) \, E^c{}_i(\underline{\Sigma}(\underline{\sigma})) \,. \tag{1.112}$$

n_a is the normal one-form

$$n_a = \epsilon_{abc} \frac{\partial f^b}{\partial \sigma^1} \frac{\partial f^c}{\partial \sigma^2} \tag{1.113}$$

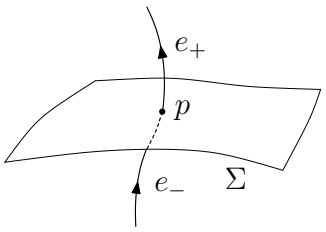


Figure 1.3: Intersection of an edge e with a surface Σ.

of Σ.

The Poisson brackets of A and E imply Poisson brackets for the holonomy and the electric flux. To determine them, we consider first the bracket between the holonomy and the unsmeared E-field: it is easy to see that

$$\left\{E_i^a(\underline{x}), U_e^j\right\} = \mathrm{i}\kappa\beta \int_0^1 \mathrm{d}s\, \dot{e}^a(s)\, \delta(\underline{x} - \underline{e}(s))\, U^j_{e_+(\underline{x})}\, {}^jJ_i\, U^j_{e_-(\underline{x})}\,. \tag{1.114}$$

$e_+(\underline{x})$ and $e_-(\underline{x})$ are the edges which result from splitting the edge e at the point $\underline{x}$. (When $\underline{x}$ is not on e, the definition of $e_+(\underline{x})$ and $e_-(\underline{x})$ is irrelevant, since the delta function gives zero.) When we include the surface integration, we get

$$\left\{E_i(\Sigma), U_e^j\right\} = \mathrm{i}\kappa\beta \int_\Sigma \mathrm{d}^2\sigma \int_0^1 \mathrm{d}s\, \frac{\partial f^a(\underline{\sigma})}{\partial\sigma^1}\frac{\partial f^b(\underline{\sigma})}{\partial\sigma^2}\frac{\mathrm{d}e^c(s)}{\mathrm{d}s}\, \epsilon_{abc}\, \delta(f(\underline{\sigma}) - \underline{e}(s))\, U^j_{e_+(\underline{x})}\, {}^jJ_i\, U^j_{e_-(\underline{x})}\,. \tag{1.115}$$

In the case where e intersects Σ in a single point (and non-tangentially), the result is simply

$$\left\{E_i(\Sigma), U_e^j[A]\right\} = \pm\mathrm{i}\kappa\beta\, U^j_{e_+}[A]\, {}^jJ_i\, U^j_{e_-}[A]\,. \tag{1.116}$$

e_+ and e_- stand for the half-edges that result from dividing e at the intersection point. The sign in front depends on the relative orientation of edge and surface parametrization. Similar formulas hold for multiple and tangential intersections. For $j = 1/2$, equation (1.116) becomes

$$\left\{E_i(\Sigma), \overline{A}_e\right\} = \pm\mathrm{i}\kappa\beta\, \overline{A}_{e_+} J_i\, \overline{A}_{e_-}\,. \tag{1.117}$$

In a standard Fock quantization, one would smear both A and E 3-dimensionally, take their Poisson bracket

$$\Big\{\int_M f\, A_a{}^i, \int_M f'\, E^b{}_j\Big\} = -\kappa\, \delta_a{}^b \int_M f f'\,, \tag{1.118}$$

and turn it into a commutation relation of operator valued distributions:

$$\left\{\hat{A}_a{}^i[f], \hat{E}^b{}_j[f']\right\} = -\mathrm{i}\hbar\kappa\, \delta_a{}^b\, \delta^i{}_j \int_M f f'\,. \tag{1.119}$$

In LQG, one uses instead the Poisson brackets (1.117) and (1.116), and elevates them to commutators

$$\left[\hat{E}_i(\Sigma), \hat{\overline{A}}_e\right] = \pm\hbar\kappa\beta\, \hat{\overline{A}}_{e_+} J_i\, \hat{\overline{A}}_{e_-}\,, \tag{1.120}$$

$$\left[\hat{E}_i(\Sigma), \hat{U}_e^j\right] = \pm\hbar\kappa\beta\, \hat{U}^j_{e_+}\, {}^jJ_i\, \hat{U}^j_{e_-}\,. \tag{1.121}$$

The smearing of E is 2-dimensional, and the exponentiation (1.105) of A can be viewed as a 1-dimensional smearing. On the right-hand side of (1.115) this is just enough to balance the delta function and give a number.

1.4.3 Kinematic Hilbert space

We go on to describe the kinematic Hilbert space on which the flux and holonomy operators are represented. The representation is of a Schrödinger type, so states are functionals over a configuration space. Since we view holonomies as the fundamental entities, we describe connections by the map $\overline{A}$, and choose $\overline{A}$ as the configuration variable. What we do is, in fact, to generalize the concept of connection by taking (1.106) and (1.106) as the *defining* properties of a connection, without assuming any longer that the $\overline{A}$ stems from a smooth field A. For that reason, the maps $\overline{A}$ are called *generalized* or *distributional connections*. The space of such connections provides the configuration space and we denote it by $\mathcal{A}$. Note that this concept of connection is similar to the one that is used in lattice gauge theory: there the holonomies provide the link variables and are restricted to edges of a fixed lattice. In quantum gravity, the distributional connections were introduced by Ashtekar and Isham [28].

To construct the Hilbert space, one defines, at first, a pre-Hilbert space of functions on $\mathcal{A}$, which is equipped with an inner product. Then, the completion in the inner product yields the full kinematic Hilbert space $\mathcal{H}_{\rm kin}$.

The pre-Hilbert space (called Cyl) consists of *cylindrical* functions: a cylindrical function depends only on the edge holonomies of a finite graph. That is, for any $\Psi \in {\rm Cyl}$, there is a graph[7] γ, consisting of edges $e_1, \ldots, e_{N_\gamma}$, and a smooth function

$$f : SU(2)^{N_\gamma} \to \mathbb{C}\,, \tag{1.122}$$

so that

$$\Psi[\overline{A}] = \Psi_{\gamma,f}[\overline{A}] := f(\overline{A}_{e_1}, \ldots, \overline{A}_{e_{N_\gamma}})\,. \tag{1.123}$$

The inner product on Cyl is constructed as follows: for any pair $\Psi_{\gamma,f}$ and $\Psi_{\gamma',f'}$ of cylindrical functions, we can find a new graph $\tilde{\gamma}$ that contains both γ and γ', and extend the dependence of $\Psi_{\gamma,f}$ and $\Psi_{\gamma',f'}$ trivially to all edges of $\tilde{\gamma}$. Then, the inner product is defined by the multiple integral

$$\langle \Psi^*_{\gamma,f}, \Psi_{\gamma',f'} \rangle := \int \left(\prod_{e\in\tilde{\gamma}} {\rm d}g_e \right) f^*(g_{e_1}, \ldots, g_{e_{N_{\tilde{\gamma}}}})\, f'(g_{e_1}, \ldots, g_{e_{N_{\tilde{\gamma}}}})\,. \tag{1.124}$$

By completing Cyl w.r.t. $\langle\ ,\ \rangle$, we arrive at the kinematic Hilbert space $\mathcal{H}_{\rm kin}$.

There is a particular class of cylindrical functions that can be used to define an orthonormal basis for $\mathcal{H}_{\rm kin}$. For that reason, these functions play a crucial role in the entire theory. Like any cylindrical function, they are determined by a graph γ, but, in addition, the edges of γ carry representation labels j_e and the vertices have tensor labels T_v. The labels replace the function f in (1.122) and specify the functional dependence on holonomies. The labelled graph is called *spin network* and the associated element of Cyl is the *spin network state*. Here, we denote spin networks by the letter S, and the state by Ψ_S [8].

For a given connection $\overline{A}$, the value $\Psi_S[\overline{A}]$ is obtained as follows: for every edge e of the spin network graph, there is a holonomy $\overline{A}_e$ and its j_e-representation $R_{j_e}(\overline{A}_e)$. We contract all such representation matrices with tensors from the vertices, in the way indicated by the graph, and thereby receive a number—the value of Ψ_S on the connection $\overline{A}$. In the example shown in Fig. 1.4, we get

$$\Psi_S[\overline{A}] = T_{1a_1a_2a_3}\, R_{j_1}(\overline{A}_{e_1})^{a_1}{}_{b_1}\, R_{j_2}(\overline{A}_{e_2})^{a_2}{}_{b_2}\, R_{j_3}(\overline{A}_{e_2})^{a_3}{}_{b_3}\, T_2{}^{b_1b_2b_3} \tag{1.125}$$

For a general spin network S, we write

[7] It is assumed that the edges of γ meet, if at all, at their starting or end points.

[8] The discovery of these states is due to Rovelli and Smolin [29], and was motivated by earlier ideas of Penrose [30].

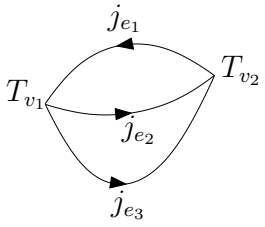


Figure 1.4: Example of a spin network.

$$\Psi_S[\overline{A}] = \left(\prod_{v\in\gamma} T_v\right) \cdot \left(\prod_{e\in\gamma} R_{j_e}(\overline{A}_e)\right) \tag{1.126}$$

where the dot $\cdot$ symbolizes the contraction of tensor indices.

>From the set of all spin network states we can choose subsets that form orthonormal bases for $\mathcal{H}_{\text{kin}}$. To prove this, one considers first a fixed graph γ, and the space Cyl_γ of cylindrical functions with support on γ. By applying the Peter-Weyl theorem to each edge of the graph, we find an orthonormal basis of spin networks for Cyl_γ. Since

$$\text{Cyl} = \bigcup_{\gamma\in M} \text{Cyl}_\gamma \tag{1.127}$$

and Cyl is dense in $\mathcal{H}_{\text{kin}}$, by construction, these bases can be patched together to give a spin network basis for the entire functional space $\mathcal{H}_{\text{kin}}$.

An important property of spin network states consists in the fact that two of them are orthogonal unless their graphs and representation labels are identical[9]. This implies immediately that a basis of spin networks is uncountable.

To complete the Schrödinger representation, we need to specify how the holonomy and flux operators act on $\mathcal{H}_{\text{kin}}$. The obvious choice for $\hat{\overline{A}}$ is

$$(\hat{\overline{A}}_e)^a{}_b\, \Psi_S[\overline{A}] := (\overline{A}_e)^a{}_b\, \Psi_S[\overline{A}]\,, \tag{1.128}$$

where $(\overline{A}_e)^a{}_b$ is a matrix entry of $\overline{A}_e \in SU(2)$. Likewise, we set

$$(\hat{U}^j_e)^a{}_b\, \Psi_S[\overline{A}] := (U^j_e)^a{}_b[\overline{A}]\, \Psi_S[\overline{A}]\,, \tag{1.129}$$

for holonomy operators of any representation j.

As concerns the flux operator $\hat{E}_i(\Sigma)$, it is enough to fix its action on spin network states, since the latter span $\mathcal{H}_{\text{kin}}$. To start with, we consider a state that corresponds to a single edge, i.e.

$$\Psi_S[\overline{A}] = (U^j_e)^a{}_b[\overline{A}]\,. \tag{1.130}$$

Then, equation (1.120) and the Leibniz rule (which we assume for $\hat{E}_i(\Sigma_i)$) imply that

$$\hat{E}_i(\Sigma_i)\, \Psi_S[\overline{A}] = \hat{E}_i(\Sigma_i)\, (U^j_e)^a{}_b[\overline{A}] = \pm\kappa\beta\, \left(U^j_{e_+}[\overline{A}]\, {}^jJ_i\, U^j_{e_-}[\overline{A}]\right)^a{}_b \tag{1.131}$$

This was for the case, when the edge hits the surface in a single point and non-tangentially. When there is no intersection, the result is zero, and there is a modified formula for tangential intersections.

[9]To be precise, we should say: "... unless the non-trivial part of the graphs is identical.", since trivial extensions of the graphs, i.e. by edges with $j = 0$, have no effect.

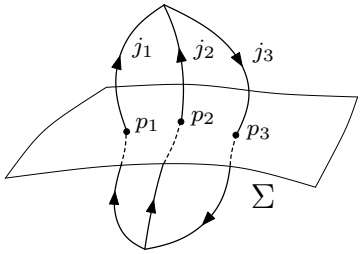


Figure 1.5: Intersection of a spin network S with a surface Σ.

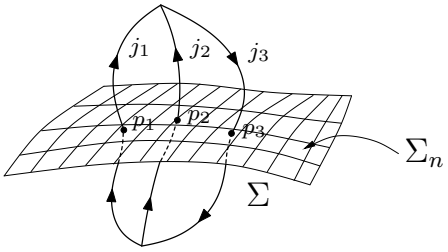


Figure 1.6: Subdivision of Σ by smaller surfaces Σ_n.

Any spin network state is just a contraction of holonomies U_e^j, so it is straightforward to extend (1.131) to a general state. When the spin network intersects the surface in several points p_k, we get a sum

$$\hat{E}^i(\Sigma)\,\Psi_S[\overline{A}] = \sum_k \pm\kappa\beta\; U^j_{e_{k+}}[\overline{A}]\;{}^{j_k}J_i\;U^j_{e_{k-}}[\overline{A}]\cdot\Psi^{(k)}_S[\overline{A}]\,, \tag{1.132}$$

where e_{k-} and e_{k+} designate the respective half-edges (see Fig. 1.5). For each k, the symbol $\cdot\Psi^{(k)}_S[\overline{A}]$ indicates the contraction with the rest of the state that was not affected by the flux operator.

1.4.4 Spin networks as quanta of geometry

To arrive at a physical interpretation of spin networks, we need to find operators that diagonalize them. In the case of a single intersection, we can square the electric flux operator in (1.132), and obtain the eigenvalue equation

$$\hat{E}^i(\Sigma)\hat{E}^i(\Sigma)\,\Psi_S[\overline{A}] = \beta^2 l_p^4\, j(j+1)\,\Psi_S[\overline{A}]\,. \tag{1.133}$$

(Note the appearance of the Planck length $l_p = \sqrt{\hbar\kappa}$!) If there are several intersections, the squaring leads to cross-terms where the J_i's are inserted on different intersection points. As a result, the diagonal action is lost. We can bypass that difficulty by considering instead the

operator

$$\sum_{n=1}^{N} \hat{E}^i(\Sigma_n)\hat{E}^i(\Sigma_n)\,, \tag{1.134}$$

where the sum runs over smaller surfaces Σ_n that subdivide Σ. If the partition is fine enough, each spin network edge e intersects each surface Σ_n only once, and the action becomes diagonal (see Fig. 1.6). To ensure that for any spin network, we take the limit of infinite partition fineness. That gives us

$$\lim_{N\to\infty} \sum_{n=1}^{N} \hat{E}^i(\Sigma_n)\hat{E}^i(\Sigma_n)\,\Psi_S[\overline{A}] = \sum_k \beta^2 l_p^4\, j_k(j_k+1)\,\Psi_S[\overline{A}]\,, \tag{1.135}$$

and also

$$\lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{\hat{E}^i(\Sigma_n)\hat{E}^i(\Sigma_n)}\,\Psi_S[\overline{A}] = \sum_k \beta l_p^2 \sqrt{j_k(j_k+1)}\,\Psi_S[\overline{A}]\,. \tag{1.136}$$

What is the physical meaning of this equation? To answer that question, we consider the classical analogue of the operator in (1.136). It can be shown that the limit of the sum gives an integral, namely,

$$\lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{E^i(\Sigma_n)E^i(\Sigma_n)} = \int_\Sigma \mathrm{d}^2\sigma\ \sqrt{n_a(\underline{\sigma})E^{ai}(f(\underline{\sigma}))\,n_b(\underline{\sigma})E^{bi}(f(\underline{\sigma}))}\,. \tag{1.137}$$

We observe furthermore that

$$n_a E^{ai} n_b E^{bi} = g\, n_a n_b\, g^{ab}\,, \tag{1.138}$$

and by the inverse formula for matrices

$$n_a n_b\, g^{ab} = \frac{\det(g_{ab} - n^{-2} n_a n_b)}{g}\,. \tag{1.139}$$

The quantity $g_{ab} - n^{-2}n_a n_b$ is the 2-metric h_{ab} that is induced on Σ by g_{ab}. Therefore,

$$n_a E^{ai} n_b E^{bi} = \det(h_{ab}) = h\,, \tag{1.140}$$

and (1.137) equals the area

$$A(\Sigma) = \int_\Sigma \mathrm{d}^2\sigma\ \sqrt{h} \tag{1.141}$$

of the surface Σ. This motivates us to define the area operator as

$$\hat{A}(\Sigma) = \sum_{n=1}^{N} \sqrt{\hat{E}^i(\Sigma_n)\hat{E}^i(\Sigma_n)} \tag{1.142}$$

Accordingly, spin network states are eigenstates of area w.r.t. the surface Σ:

$$\hat{A}(\Sigma)\,\Psi_S = \sum_k \beta l_p^2 \sqrt{j_k(j_k+1)}\,\Psi_S \tag{1.143}$$

The spectrum of eigenvalues is discrete and there is a minimum eigenvalue. Thus, spin networks may be viewed as quanta of area, in a similar way as particles are quanta of energy. By analogous methods it is also possible to introduce a volume operator. Like the area

operator, it has a discrete spectrum, but the determination of eigenstates is much more complicated.

The discreteness of geometrical spectra was first established by Rovelli and Smolin [31], and is one of the principal results of canonical LQG. It means that the theory describes spaces which have a discrete nature. Spin networks correspond to the different possible quantum states of space: we interpret vertices as "chunks" of space volume, and labelled edges correspond to quanta of area between these volumes. Intuitively, one would expect that the wavelength of any object on such a quantum space has an upper bound, since space is only constituted by the discrete graph of the spin network. This gives rise to the hope that canonical LQG has a built-in cutoff that regulates both gravity and the matter fields that are coupled to it. If this is correct, it would be an *intrinsic* cutoff in the sense that it is not put in by hand, as in effective field theories, but implied by the quantization procedure.

1.4.5 Imposition of gauge and diffeomorphism constraint

According to the Dirac program, the physical content of the theory has to be extracted by imposing invariance under gauge-transformations, diffeomorphisms and the action of the scalar constraint. Below we explain the procedure for gauge- and diff-transformations, which is relatively straightforward. The implementation of the scalar constraint, on the other hand, is a rather open and debated issue: we will briefly comment on it in the next section.

On generalized connections, gauge transformations act by

$$\overline{A} \mapsto \overline{A}^{\lambda}\,, \qquad \overline{A}^{\lambda}_e := \lambda_{e(1)} \overline{A}_e \lambda_{e(1)}\,, \tag{1.144}$$

where λ is a function

$$\lambda : M \to SU(2)\,. \tag{1.145}$$

$e(0)$ and $e(1)$ denote the starting and end point of an edge. This induces a natural action $\Psi \mapsto \Psi^{\lambda}$ on states, namely,

$$\Psi^{\lambda}[\overline{A}] := \Psi[\overline{A}^{\lambda^{-1}}]\,. \tag{1.146}$$

On a spin network state such a transformation has the effect of transforming tensors at vertices in their respective representation: e.g.

$$T_v{}^{a_1}{}_{a_2 a_3}{}^{a_4}{}_{a_5} \quad \in \quad V_{j_1} \otimes V^*_{j_2} \otimes V^*_{j_3} \otimes V_{j_4} \otimes V^*_{j_4} \tag{1.147}$$

transforms to

$$R_{j_1}{}^{a_1}{}_{b_1}(\lambda_v)\, R^*_{j_2 a_2}{}^{b_2}(\lambda_v)\, R^*_{j_3 a_3}{}^{b_3}(\lambda_v)\, R_{j_4}{}^{a_4}{}_{b_4}(\lambda_v)\, R^*_{j_5 a_5}{}^{b_5}(\lambda_v)\, T_v{}^{b_1}{}_{b_2 b_3}{}^{b_4}{}_{b_5}\,. \tag{1.148}$$

Thus, the solution of the gauge-constraint on $\mathcal{H}_{\rm kin}$ is obvious: gauge-invariant spin network states are those which carry invariant tensors I_v as labels. The gauge-invariant Hilbert space $\mathcal{H}_0$ is the span of the gauge-invariant spin networks.

The situation is slightly more subtle when we come to invariance under spatial diffeomorphisms. For a diffeomorphism $f : M \to M$, there is a natural action on generalized connections, given by

$$\overline{A} \mapsto \overline{A}^{f}\,, \qquad \overline{A}^{f}_e := \overline{A}_{f\circ e}\,. \tag{1.149}$$

and that induces the action

$$\Psi^{f}[\overline{A}] := \Psi[\overline{A}^{f^{-1}}] \tag{1.150}$$

on states. On spin network states, this action has a very simple effect: it maps the graph γ to the new graph $f_*\gamma$—i.e. by mapping edges and vertices with f—and leaves the labels unchanged (see Fig. 1.7). It is this simple behaviour under diffeomorphims that makes the

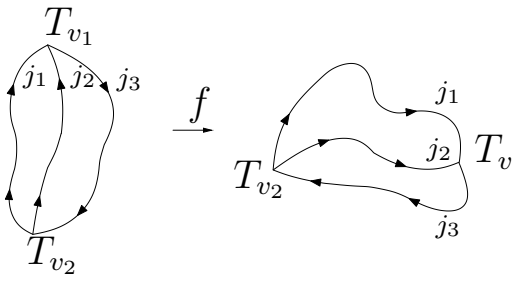


Figure 1.7: Action of a diffeomorphism f on a spin network.

concept of spin networks so appealing in a background-independent theory. We can almost immediately identify which features of the spin network encode diff-invariant information, and which parts of it are diff-dependent—and therefore gauge. The invariant part is, roughly speaking, the labelling, the connectivity and the knotting[10], while the gauge part corresponds to the choice of embeddings that preserve these properties. Note that, in comparison, it would be much harder to identify the diff-invariant properties of a metric field.

Having said so, how should we implement the invariance condition on the Hilbert space? If we followed strictly the Dirac program, we would impose

$$\hat{V}_a \Psi_S = 0 \tag{1.151}$$

i.e. invariance under infinitesimal spatial diffeomorphisms. In the present case, that does not make any sense, since the action of diffeomorphisms on spin network states is not continuous—if two spin networks differ only slightly in their graphs, they are orthogonal.

The next possibility would be to demand invariance under finite diffeomorphisms, i.e.

$$\Psi_S^f = \Psi_S \tag{1.152}$$

for all diffeomorphisms $f : M \to M$. Now, the condition is well-defined, but it has no solution in $\mathcal{H}_0$, except for the empty spin network: there is no non-trivial spin network state that is invariant under the action of all diffeomorphisms.

The resolution of this difficulty lies in what we said before about the diff-invariant properties of spin networks. When applying Dirac's rules, we have to be aware that they were formulated for mechanics and field theories. If we interpret them too literally, and use them without adaptation in other contexts, they might lead to meaningless results. What we should rather do is to concentrate on their physical meaning, namely, the requirement that the unphysical degrees of freedom be eliminated from the states. In a standard field theory, equation (1.151) is the correct way to implement this condition. Here, however, the theory is no longer formulated in terms of smooth fields, but instead with generalized connections. In that context, neither (1.151) nor (1.152) allow us to eliminate the diff-dependent properties of the states.

A method that *does* achieve this is the following: we replace the concrete, embedded spin networks, by equivalence classes of spin networks under diffeomorphisms. These are called *abstract spin networks* s. By using equivalence classes, we discard all diff-dependent information in the spin networks, and retain only the invariant features we mentioned above. To each abstract spin network s, we associate a state Ψ_s, and define the diff-invariant Hilbert

[10]We could unknot edges, if we were allowed to act on each edge separately, but here the diffeomporphism acts on all edges at once.

space $\mathcal{H}_{\text{diff}}$ as the space of formal (and finite) linear combinations of Ψ_s. We specify the inner product by

$$\langle \Psi_s, \Psi_{s'} \rangle = \sum_{S \in s} \langle \Psi_S, \Psi_{S'} \rangle \,, \qquad S' \in s' \,. \tag{1.153}$$

Here, S' is an arbitrary representative of s', and the sum extends over all representatives S of s. Due to the orthogonality properties of spin networks, $\langle \Psi_S, \Psi_{S'} \rangle$ is only non-zero when the graph and spin labelling of S and S' are congruent. Thus, the sum counts all ways of embedding s on S', and adds up the inner products. There are only a finite number of possibilities, so the result is well-defined.

The same construction can be also understood in terms of a Gelfand triple and a group-averaging method. We consider the Gelfand triple

$$\text{Cyl} \subset \mathcal{H}_{\text{diff}} \subset \text{Cyl}^* \tag{1.154}$$

and seek solutions to (1.152) in Cyl^*. That is, we look for dual states $\Phi \in \text{Cyl}^*$ that have the property

$$\Phi(\Psi_S^f) = \Phi(\Psi) \,. \tag{1.155}$$

Such solutions can be easily found by an averaging over the diffeomorphism group. For a given spin network state Ψ_S, we define a diff-averaged dual state

$$\Phi_S := \sum_{g \in \text{Diff}(M)} \Psi_S^{g*} \,. \tag{1.156}$$

Here, Ψ_S^* denotes the element in Cyl^* that is dual to Ψ_S. The right-hand side of (1.156) is well-defined, since upon contraction with a state $\Psi_{S'}$ only a finite number of terms survive. The state Φ_S is clearly invariant under diffeomorphisms, since

$$\Phi_S^f = \sum_{f \in \text{Diff}(M)} \Psi_S^{gf*} = \Phi_S \,. \tag{1.157}$$

Moreover, Φ_S depends only on the diff-equivalence class s of S, so we write $\Phi_s \equiv \Phi_S$. In this approach, we define $\mathcal{H}_{\text{diff}}$ as the space spanned by the Φ_s, where the inner product is given by

$$\langle \Phi_s, \Phi_{s'} \rangle = \Phi_s(\Psi_{S'}) \,, \qquad S' \in s \,. \tag{1.158}$$

We see that this is equivalent to (1.153), since

$$\begin{aligned} \langle \Phi_s, \Phi_{s'} \rangle &= \sum_{f \in \text{Diff}(M)} \Psi_S^{f*}(\Psi_{S'}) && (1.159) \\ &= \sum_{f \in \text{Diff}(M)} \langle \Psi_S^f, \Psi_{S'} \rangle && (1.160) \\ &= \sum_{S'' \in s} \langle \Psi_{S''}, \Psi_{S'} \rangle \,. && (1.161) \end{aligned}$$

1.4.6 Dynamics

We finally make some remarks on the scalar constraint and the treatment of dynamics. Ideally, the aim would be to define a scalar constraint operator $\hat{C}$ or a finite evolution map

$$\exp\left(\frac{\mathrm{i}}{\hbar} \int_M N \hat{C} \right) \tag{1.162}$$

on the gauge-invariant Hilbert space $\mathcal{H}_0$ (1), to translate it into an operator on the diff-invariant Hilbert space $\mathcal{H}_{\rm diff}$ (2), and use it for determining states of the physical Hilbert space $\mathcal{H}_{\rm phys}$ (3). Transition amplitudes would be encoded by an—as yet to be defined—inner product on $\mathcal{H}_{\rm phys}$ (4).

At present, all four points are open problems. So far, the most prominent work on dynamics is Thomas Thiemann's proposal for a scalar constraint operator ([32]–[33]). We briefly sketch the logic of its construction.

Let Ψ_S be a spin network state $\mathcal{H}_0$, and choose a triangualation κ_ϵ, so that all edges of the spin network lie on edges of the triangulation. ϵ parametrizes the triangulation fineness. Thiemann translates the complicated classical constraint C (see (1.99)) into an operator $\hat{C}_\epsilon$ that acts on Ψ_S. It has, roughly speaking, the effect of adding edges around four-valent vertices of the spin network. When ϵ is small enough, a further decrease in ϵ has only the effect of moving the new edge closer to the vertex, but everything else remains unchanged. To deal with the residual ϵ-dependence, one takes the dual $\hat{C}^*_\epsilon$ of the operator w.r.t. $\mathcal{H}_{\rm diff} \subset {\rm Cyl}^*$, and the limit

$$C^* := \lim_{\epsilon\to 0} \hat{C}^*_\epsilon \,. \tag{1.163}$$

The latter is well-defined, since the shifting of the edge towards the vertex corresponds to a diffeomorphism, and dual states are insensitive to that.

There are a number of critical issues about this construction:

- It is not clear if the infinitesimal form of the constraint is appropriate for implementing the quantum dynamics.
- The operator C^* is not an operator on $\mathcal{H}_0$, and it is not clear if it can be interpreted as the scalar constraint on $\mathcal{H}_{\rm diff}$. If one found a definition of $\hat{C}$ on $\mathcal{H}_0$, the projection to $\mathcal{H}_{\rm diff}$ would involve a diff-averaging, and formally

  $$\langle \Psi_s, \hat{C}_{\text{on } \mathcal{H}_{\rm diff}} \Psi_{s'} \rangle = \frac{1}{V({\rm Diff}(M)} \sum_{f,g\in {\rm Diff}(M)} \langle \Psi^f_S,\, \hat{C}\, \Psi^g_{S'} \rangle \tag{1.164}$$

  This is not the same as taking the dual operator, which just gives

  $$\langle C^* \Psi_s, \Psi_{s'} \rangle = \sum_{f\in {\rm Diff}(M)} \langle \Psi^f_S,\, \hat{C}\, \Psi_{S'} \rangle \,. \tag{1.165}$$

  Thus, there seems to be one diff-averaging too less in (1.163).
- The translation from classical to quantum constraint involves a triangulation that is finer than the spin network graph. If we take the viewpoint that spin networks constitute space, it would appear more natural to identify triangulation and graph.
- The last point is related to the problem that the operator $\hat{C}^*$ is ultralocal: it creates only links between edges of the same vertex. This property could imply a lack of long-range correlations in physial solutions [34].

Recently, Thiemann has proposed a new technique for implementing the scalar constraint that could circumvent many problems of his original proposal [35]–[36].

1.5 Spin foam gravity

The spin foam approach aims at the construction of rigorously defined covariant models of quantum gravity. By "covariant" we mean that the description is based on a sum-over-histories picture, as opposed to the state-evolution picture of the Hamiltonian formalism.

As in canonical LQG, a central role is played by the connection, and states are given by spin network states, or their generalizations to gauge groups other than SU(2). "Histories" of the system are represented by labelled 2-complexes, called spin foams, that interpolate between spin networks in the boundary.

In the following, we present the two most known models: the spin foam model of 3d gravity (sec. 1.5.1 and 1.5.2), and the Barrett-Crane model for 4d gravity (sec. 1.5.3 and 1.5.4). For simplicity, we consider only Riemannian gravity.

In the 3-dimensional case, one can prove the equivalence of spin foam and canonical loop quantization, while in 4 dimensions the relation between the two approaches is far less understood. Rovelli and Reisenberger have given a general argument which indicates that transition amplitudes of canonical LQG equal a sum over spin foams [37]. The idea is to expand the matrix element of a finite evolution as

$$\langle \Psi_{S_\mathrm{f}}| \exp\left(\frac{\mathrm{i}}{\hbar}\int_M N\hat{C}\right)|\Psi_{S_\mathrm{i}}\rangle = \sum_{n=0}^{\infty}\langle \Psi_{S_\mathrm{f}}|\left(\frac{\mathrm{i}}{\hbar}\int_M N\hat{C}\right)^n|\Psi_{S_\mathrm{i}}\rangle\,. \tag{1.166}$$

In each term, the scalar constraint maps the initial spin network into a superposition of new spin networks, and the repeated action of $\hat{C}$ yields sequences of spin networks. Each sequence can be viewed as a spin foam whose faces carry the spin labels of the spin network edges and whose edges carry the intertwiner labels of the spin network vertices. Thus, the amplitude (1.166) is reexpressed as a sum over spin foams.

The argument supports the idea that there should be a spin foam formulation of quantized Ashtekar-Barbero gravity. However, since the implementation of dynamics is not clear in that case, there is no concrete proposal for an associated spin foam sum. It seems that it cannot correspond to any of the existing spin foam models of Lorentzian gravity, since these are labelled by irreps of the Lorentz group, and not by SU(2) irreps.

1.5.1 3d gravity as a BF theory

The Palatini formulation of 3d gravity has the Lagrangian

$$\mathcal{L} = \epsilon^{\alpha\beta\gamma}\epsilon_{IJK}\, e_\alpha{}^I\, F_{\beta\gamma}{}^{JK}\,. \tag{1.167}$$

The variational principle yields

$$F_{\beta\gamma}{}^{JK} = 0\,, \tag{1.168}$$

$$D_\beta\left(\epsilon^{\alpha\beta\gamma}\epsilon_{IJK}\, e_\alpha{}^I\right) = 0\,. \tag{1.169}$$

The second equation implies that

$$D_{[\alpha}e_{\beta]}{}^I = 0 \tag{1.170}$$

and therefore A is compatible with the triad (in our notation $A = \omega$). That means that $F_{\beta\gamma}{}^{JK} = R_{\beta\gamma}{}^{JK} = 0$, which is equivalent to Einstein's equations in 3 dimensions.

After a Legendre transform, we obtain the Poisson brackets

$$\left\{A_a{}^I, E^b{}_J\right\} = \kappa\,\delta_a{}^b\delta^I{}_J\delta(\underline{x}-\underline{y}) \tag{1.171}$$

where

$$A_a{}^I := \epsilon^{IJK}A_{\alpha JK}\,, \qquad \text{and} \qquad E^{aI} := \sqrt{g}\,e^{aI}\,. \tag{1.172}$$

The constraints have the same form as for self-dual gravity in 4 dimensions[11]:

$$G_I = D_a E^a{}_I \tag{1.173}$$
$$V_a = E^b{}_I F_{ab}{}^I \tag{1.174}$$
$$C = \epsilon_{IJK} E^{aI} E^{bJ} F_{ab}{}^K , \tag{1.175}$$

Let us count the degrees from freedom: we have 6 for the phase space (6 $\omega_a{}^{IJ}$, 6 e_aI), 3+2+1 first-class constraints, so there appear no second-class constraints. At least locally, the number of physical degrees of freedom is 0. (Our infinitesimal analysis can only reflect local properties, so there remains the possibility of global degrees of freedom.)

The gravity Lagrangian (1.167) can be seen as a special case of a more general Lagrangian, which is defined for arbitrary gauge group G and dimension d. The associated theory is called BF theory and has the action

$$S = \int \mathrm{tr}\,[B \wedge F(A)] \tag{1.176}$$

The variables are a connection one-form A and a Lie algebra-valued $(d-2)$-form field B. The trace symbol stands for the trace in the fundamental matrix representation of the Lie algebra elements.

In the case of $d = 3$ and $G = \mathrm{SU}(2)$, the BF action is equivalent to that of 3d Riemannian gravity, as can be seen as follows: we expand the Lie algebra-valued forms in the basis $\mathrm{i}\sigma_I/2$, $I = 1, 2, 3$

$$B = B^I \,\mathrm{i}\sigma_I/2\,, \qquad \text{and} \qquad F = F^I \,\mathrm{i}\sigma_I/2\,, \tag{1.177}$$

where σ_I, $I = 1, 2, 3$, are the Pauli matrices satisfying

$$\mathrm{tr}(\sigma_I \sigma_J) = 2\,\delta_{IJ}\,. \tag{1.178}$$

Then,

$$\mathrm{tr}\,[B \wedge F(\omega)] = B^I \wedge F^J \,\mathrm{tr}(\mathrm{i}\sigma_I/2\,\mathrm{i}\sigma_J/2) \tag{1.179}$$
$$= -\frac{1}{2}\, B^I \wedge F_I\,. \tag{1.180}$$

The Lie algebras of SU(2) and SO(3) are isomorphic under the map (of basis elements)

$$\mathrm{su}(2) \ni \mathrm{i}\sigma_I/2 \quad \mapsto \quad J_I \in \mathrm{so}(3)\,, \qquad I = 1, 2, 3, \tag{1.181}$$

where

$$(J_I)^K{}_L := \epsilon_I{}^K{}_L\,. \tag{1.182}$$

Thus, we can interpret the forms A^I and F^I also as components of so(3)-valued forms A and F. In that case, we get

$$F^K{}_L = F^I (J_I)^K{}_L = F^I\, \epsilon_I{}^K{}_L\,, \tag{1.183}$$

or

$$F^I = \frac{1}{2} \epsilon^I{}_{JK}\, F^{JK}\,. \tag{1.184}$$

By inserting this in (1.180), and identifying B^I with the co-triad e^I, we get the gravity Lagrangian:

$$-\frac{1}{4}\,\epsilon_{IJK}\, e^I \wedge F^{JK}\,, \tag{1.185}$$

[11] Note that the indices $I, J, \ldots$ go from 1 to 3.

1.5.2 Spin foam model of 3d gravity

There are essentially two methods for constructing the spin foam sum of 3d gravity. One of them departs from a canonical quantization, and the sum over spin foams results from the imposition of the scalar constraint [38]. The second method is covariant in the sense that it does not involve a 2+1 decomposition: it starts from the BF-formulation (with $d = 3$, $G =$SU(2)), and reaches the spin foam sum by means of a dual transformation[12].

Here, we will sketch the second of the two methods. Note that the same construction exists also for Lorentzian 3d gravity, i.e. for gauge group SO(2,1).

We begin by replacing the continuous 3-manifold by a triangulation κ. We also choose a complex κ^* that is dual to κ. In the following, only the 0-, 1- and 2-cells of κ^* will be relevant. Thus, we may remove the higher-dimensional cells and think of κ^* as a 2-complex. The one-form B is translated into a Lie algebra element B_l for each edge[13] l of the triangulation κ. As in canonical LQG, the connection A is replaced by holonomies—group elements g_{l^*} that are associated to each dual edge l^* of the dual complex κ^*. Since edges are in one-to-one correspondence to dual faces f^*, we can write the B-variable also as B_{f^*}. The BF action (1.176) is represented as

$$S = \sum_{f^* \in \kappa^*} \mathrm{tr}\left(B_{f^*} U_{f^*}\right) , \tag{1.186}$$

where U_{f^*} denotes the product $g_{l_n^*} \cdots g_{l_1^*}$ of group elements around the dual face f^*. The rational behind this is the naive continuum limit: for

$$U_{f^*} \equiv \mathbb{1} + F_{f^*} \approx \mathbb{1} \tag{1.187}$$

we can find smooth fields B and A such that

$$\sum_{f^* \in \kappa^*} \mathrm{tr}\left(B_{f^*} U_{f^*}\right) = \sum_{f^* \in \kappa^*} \left[\underbrace{\mathrm{tr}\left(B_{f^*} \mathbb{1}\right)}_{=0} + \mathrm{tr}\left(B_{f^*} F_{f^*}\right) \right] \tag{1.188}$$

$$\approx \int_M \mathrm{tr}\left(B \wedge F\right) . \tag{1.189}$$

The partition function of the quantum theory is defined by the path integral

$$Z(\kappa) = \int \left(\prod_{f^* \in \kappa^*} \mathrm{d}B_{f^*} \right) \left(\prod_{l^* \in \kappa^*} \mathrm{d}g_{l^*} \right) \exp\left(\mathrm{i} \sum_{f^* \in \kappa^*} \mathrm{tr}[B_{f^*} U_{f^*}] \right) \tag{1.190}$$

The measure on g_{l^*} is the Haar measure SU(2), and $B_{f^*} = B^I_{f^*}\, \mathrm{i}\sigma_I/2$ is integrated with the Lebesgue measure on $\mathbb{R}^3$. It can be shown that the integration over each B_{f^*} yields a delta function of the holonomy U_{f^*}, i.e.

$$Z(\kappa) = \int \left(\prod_{l^* \in \kappa^*} \mathrm{d}g_{l^*} \right) \prod_{f^* \in \kappa^*} \delta(U_{f^*}) \tag{1.191}$$

At this point, the theory has the form of a pure lattice gauge theory, and we can apply the dual transformation to the spin foam model: by that we mean that we expand the amplitude for each dual face in characters, and subsequently integrate over group variables.

[12] For any pure gauge theory, there exists a dual transformation to a spin foam model, which we explain in more detail in chapter 4.

[13] In this section, we use the letter l for edges to avoid confusion with the triad e.

The character expansion is[14]:

$$\delta(U_{f^*}) = \sum_{j \in \mathbb{N}_0/2} (2j+1)\, \chi_j(U_{f^*})\,, \tag{1.192}$$

and by inserting this in (1.191) we get

$$Z(\kappa) = \int \left(\prod_{l^* \in \kappa^*} \mathrm{d}g_{l^*} \right) \sum_{\{j_{f^*}\}} \prod_{f^* \in \kappa^*} (2j_{f^*}+1)\, \chi_j(U_{f^*})\,. \tag{1.193}$$

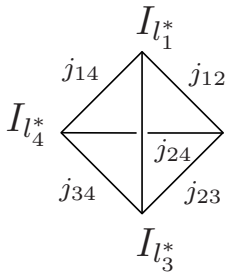
The holonomy around a dual face is a product $g_{l_n^*} \cdots g_{l_1^*}$, so we can rewrite the character as a contraction of representation matrices of the $g_{l_i^*}$:

$$\chi_j(U_{f^*}) = R_j(g_{l_n^*}) \cdots R_j(g_{l_1^*}) \tag{1.194}$$

The next step is to interchange the order of integral and sum, and to integrate out the group variables. Here, we only state the result, as it is explained in more detail in chapter 4. The integration produces a sum over assignments of invariant tensors I_{l^*} to dual edges, and an amplitude factor for each vertex of the dual complex (or tetrahedron of the triangulation):

$$Z(\kappa) = \sum_{\{j_{f^*}\}} \sum_{\{I_{l^*}\}} \prod_{f^* \in \kappa^*} (2j_{f^*}+1) \prod_{v^* \in \kappa^*} A_{v^*}\,. \tag{1.195}$$

The admissible tensors I_{l^*} come from an orthonormal basis of invariant tensors in $V_{j_1} \otimes V_{j_2} \otimes \ldots$, where $j_1, j_2, \ldots$ are the labels of the dual faces incident on l^*. The vertex amplitude results from contracting all invariant tensors from edges that run into the vertex: we symbolize this by

$A_{v^*} =$ 

. (1.196)

Here, the j_{ij}'s denote the labels of the dual faces that are bounded by pairs of edges l_i^*, l_j^*. In the present case ($d = 3$, $G =$SU(2)), each dual edge has only three incident dual faces, and the invariant tensor is uniquely fixed by the labels. The sum over I_{l^*} in (1.195) drops out, and (1.196) is simply the 6j-symbol.

Each term in (1.195) corresponds to a labelling of the 2-complex κ^* with irreducible representations and invariant tensors. We call such a labelled 2-complex a *spin foam*, and regard (1.195) as a sum over spin foams. In general, spin foams may have different underlying 2-complexes, but in the sum (1.195) the complex is fixed to be κ^*.

More generally, the transition from path integral to spin foam sum can be also done for transition amplitudes, i.e. for path integrals that are weighted with boundary functionals. As in canonical LQG, the states translate into superpositions of spin networks, and the admissible spin foams in the sum are those that interpolate between the boundary spin networks.

We should mention that there appear divergences in the BF spin foam sum, which can be interpreted as a consequence of a symmetry in the original BF action. There is a prescription

[14]Strictly speaking, the integration over B gives $\tilde{\delta}(U) = \sum_{j \in \mathbb{N}_0} (2j+1)\, \chi_j(U_{f^*})$, where the sum runs only over *even* j (see Appendix B in [39]). It is customary, however, to include also half-integer representations.

for dividing out the infinite gauge volume associated to this symmetry, which makes transition amplitudes mathematically well-defined [40]. The resulting "renormalized" partition function $Z_r(\kappa)$ is still divergent, but independent of the original choice of triangulation κ, i.e.

$$Z_r(\kappa) = Z_r(\kappa') \tag{1.197}$$

In that sense, Z_r is a topological invariant.

The above technique provides a rigorous quantization of 3d gravity, and it can be shown to be equivalent with the quantization via Chern-Simons theory [41]. Furthermore, it is possible to include fermionic matter in the model and describe scattering between spinning massive particles. This has been done in a first-quantized approach at first [42], and later been extended to a combined perturbative quantization of scalar field theory and 3d gravity[15] [43]. The associated amplitudes equal that of an effective field theory without gravity that lives on a non-commutative spacetime.

1.5.3 4d gravity as a constrained BF theory

When we come to 4-dimensional gravity, the situation is expected to be profoundly different: now the theory has local degrees of freedom. The action can no longer be written in the BF form, and the dual transformation cannot be applied directly, since it would require the integration over the tetrad fields.

What we can do, however, is to reexpress the gravity action by a BF action with additional constraints. The strategy is to start from this constrained BF action, and construct a modified BF spin foam model from it.

We begin by discussing the classical theory. In the next subsection, we sketch the transition to the quantum theory and spin foams, which yields the so-called Barret-Crane model. The latter is the most studied 4d model in the literature, but there exist also other proposals. Here, we restrict the discussion to the Barret-Crane model, and follow the derivation described in [44]. The signature is positive, i.e. $G = \mathrm{SO}(4)$.

The idea of constraining BF theory originates from the Plebanski formulation of gravity: it has the action

$$S = \int B^{IJ} \wedge F_{IJ}(A) - \frac{1}{2}\,\phi_{IJKL} B^{IJ} \wedge B^{KL} \tag{1.198}$$

where the Lagrange multiplier ϕ satisfies $\phi_{IJKL} = -\phi_{JIKL} = -\phi_{IJLK} = \phi_{KLIJ}$ and

$$\epsilon^{IJKL}\phi_{IJKL} = 0\,. \tag{1.199}$$

The equations of motion read

$$DB = 0\,, \qquad \text{(var. of } \omega) \tag{1.200}$$

$$F_{IJ}(A) = \phi^{IJKL} B_{KL}\,, \qquad \text{(var. of } B) \tag{1.201}$$

$$B^{IJ} \wedge B^{KL} = e\,\epsilon^{IJKL}\,, \qquad \text{(var. of } \phi) \tag{1.202}$$

where we abbreviated

$$e := \frac{1}{4!}\epsilon_{IJKL}\, B^{IJ} \wedge B^{KL}\,. \tag{1.203}$$

For configurations that have $e \neq 0$, equation (1.202) is equivalent to

$$C := \epsilon_{IJKL}\, B^{IJ}_{\alpha\beta}\, B^{KL}_{\gamma\delta} - \tilde{e}\,\epsilon_{\alpha\beta\gamma\delta} = 0 \tag{1.204}$$

[15] One expands w.r.t. the coupling between scalar field and gravity. The treatment of the pure gravity term remains non-perturbative.

with

$$\tilde{e} := \frac{1}{4!}\epsilon_{IJKL}\, B^{IJ}_{\alpha\beta} B^{KL}_{\gamma\delta}\, \epsilon^{\alpha\beta\gamma\delta}\,. \tag{1.205}$$

One can show that the solutions to this equation are B's that are formed from co-triad fields e^I. More precisely, there are two sectors of solutions, namely,

$$\text{(I)} \qquad B^{IJ} = \pm\, \epsilon^{IJKL} e_K \wedge e_L\,, \tag{1.206}$$

and

$$\text{(II)} \qquad B^{IJ} = \pm\, e^I \wedge e^J\,. \tag{1.207}$$

By plugging (I) back into (1.198), together with the "solution" $\mu = 0$, we arrive at the Hilbert-Palatini action

$$S = \int \epsilon_{IJKL}\, e^I \wedge e^J \wedge F^{KL}(A)\,. \tag{1.208}$$

This tells us that the Plebanski action admits three sectors of solutions,

$$e = 0\,, \tag{1.209}$$
$$e \neq 0\,, \quad (I)\,, \tag{1.210}$$
$$e \neq 0\,, \quad (II)\,, \tag{1.211}$$

and the second one corresponds to gravity. Clearly, the last two sectors could be also obtained by using instead the BF action

$$S = \int B^{IJ} \wedge F_{IJ}(A) \tag{1.212}$$

with the additional constraints

$$e \neq 0\,, \qquad \text{and} \qquad C = 0\,. \tag{1.213}$$

1.5.4 Barrett-Crane model

In conjunction with the constraints (1.213), the BF action has gravity among its solutions. That motivates us to write down the following formal path integral:

$$\int DA \int DB\ \ \delta(C) \exp\left(\mathrm{i} \int \mathrm{tr}\left[B \wedge F(\omega)\right]\right) \tag{1.214}$$

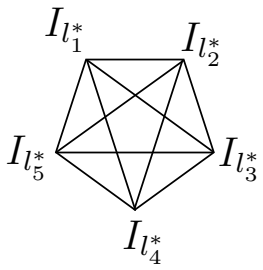
The $C = 0$ constraint is imposed by the delta function, while we ignore the $e \neq 0$ constraint. Without the constraint C, this would be a pure BF path integral, for which we know how to construct the spin foam sum: we introduce a discretized path integral

$$Z(\kappa) = \int \left(\prod_{f\in\kappa} \mathrm{d}B_f\right) \left(\prod_{l^*\in\kappa^*} \mathrm{d}g_{l^*}\right)\ \exp\left(\mathrm{i} \sum_{f\in\kappa} \mathrm{tr}[B_f U_{f^*}]\right)\,, \tag{1.215}$$

and apply the dual transformation to the spin foam sum: for SO(4) this gives

$$Z(\kappa) = \sum_{\{\rho_{f^*}\}} \sum_{\{I_{l^*}\}} \prod_{f^*\in\kappa^*} \dim \rho_{f^*} \prod_{v^*\in\kappa^*} A_{v^*}\,, \tag{1.216}$$

where

$$A_{v^*} = \quad \text{  } \quad . \tag{1.217}$$

The sum extends over all irreducible representations ρ of SO(4). For later, we note that every irrep of SO(4) is isomorphic to a product $j_l \otimes j_r$ of representations j_l, j_r of SU(2). In (1.215), the B are Lie algebra variables on faces of the triangulation κ. Since $d = 4$, there is a one-to-one correspondence between faces f and dual faces f^*.

The question is: can the spin foam sum (1.216) be modified in such a way that it corresponds to the constrained path integral (1.214)? How should the constraint (1.204) on B be implemented in the discretized path integral (1.215), and, then, at the level of the spin foam sum (1.216)?

In a first step, we translate the constraint into a condition on the B_f in (1.214). It can be shown that the constraint

$$\epsilon_{IJKL}\, B^{IJ}_{\alpha\beta}\, B^{KL}_{\gamma\delta} = \tilde{e}\, \epsilon_{\alpha\beta\gamma\delta} \tag{1.218}$$

is trivially satisfied when we evaluate both sides on vectors $v_1, \ldots, v_4$ that span $\mathbb{R}^4$. The non-trivial conditions arise when the inserted four vectors are linearly dependent. In terms of the B_f this means that the non-trivial conditions come from variables B_{f_1}, B_{f_2} where f_1 and f_2 belong to the same 4-simplex and share at least one edge. In that case, condition (1.218) translates into

$$\epsilon_{IJKL}\, B^{IJ}_{f_1}\, B^{KL}_{f_2} = 0\,, \tag{1.219}$$

since the right-hand side of (1.218) gives zero.

Next we transform (1.219) to a constraint for the spin foam sum (1.216), using a heuristic argument from [44]: consider the left-invariant vector field

$$X_V(U) := U^I{}_K V^K{}_J \frac{\partial}{\partial U^I{}_J}\,, \qquad U \in \mathrm{SO}(4)\,, \tag{1.220}$$

that is associated to the Lie algebra element $V \in \mathrm{so}(4)$. At $U = U_{f^*}$, this left-invariant vector field has the following effect on the exponentiated action:

$$\Big(-\mathrm{i} X_V(U_{f^*}) \Big) \exp\left(\mathrm{i} \sum_{f\in\kappa} \mathrm{tr}[B_f U_{f^*}] \right) \tag{1.221}$$

$$= \mathrm{tr}\left[B_f\, V\, U_{f^*}\right] \exp\left(\mathrm{i} \sum_{f\in\kappa} \mathrm{tr}[B_f U_{f^*}] \right) \tag{1.222}$$

Thus, at least for $U_{f^*} \approx \mathbb{1}$, the action of $V = J^{IJ}$ (where $(J_{IJ})^{KL} = \delta^K{}_{[I}\delta^L{}_{J]}/2$) produces

$$B^{IJ}_f \exp\left(\mathrm{i} \sum_{f\in\kappa} \mathrm{tr}[B_f U_{f^*}] \right)\,. \tag{1.223}$$

Motivated by this, we rewrite the constraint (1.219) as

$$\epsilon_{IJKL}\, X_{J_{IJ}}(U_{f^*}) X_{J_{KL}}(U_{f^*}) = 0\,, \tag{1.224}$$

and replace (1.215) by

$$Z(\kappa) = \int \left(\prod_{l^* \in \kappa^*} \mathrm{d}g_{l^*} \right) \prod_{f^* \in \kappa^*} \delta\Big(\epsilon_{IJKL}\, X_{J_{IJ}}(U_{f^*}) X_{J_{KL}}(U_{f^*}) \Big)\, \delta(U_{f^*})\,. \tag{1.225}$$

The constraint (1.224) is a Casimir of SO(4), and when going over to the spin foam sum, it is transformed into the condition

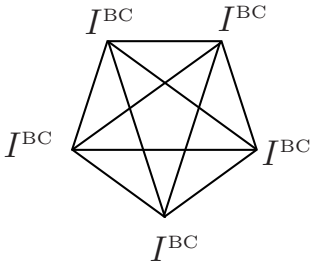
$$j_l(j_l+1) - j_r(j_r+1) = 0 \tag{1.226}$$

on irrep labels of spin foams and into a further condition on invariant tensors I_{l^*}. The condition (1.226) implies that the irreps have to satisfy $j_l = j_r$ (such irreps of SO(4) are called balanced). The condition on intertwiners requires them to be so-called Barrett-Crane intertwiners I^{BC}, whose technical specification we will not give here.

The spin foam sum becomes

$$Z(\kappa) = \sum_{\{\rho^{\mathsf{bal}}_{f^*}\}} \prod_{f^* \in \kappa^*} \dim \rho^{\mathsf{bal}}_{f^*} \prod_{v^* \in \kappa^*} A_{v^*}\,, \tag{1.227}$$

where

$A_{v^*} =$ 

. (1.228)

The Barrett-Crane intertwiners are completely fixed by incoming and outgoing irreps, so there is no sum over intertwiners. Equation (1.227) defines the partition function of the Barrett-Crane model. Transition amplitudes are constructed analogously.

1.6 Contents of the thesis

Based on the above introduction, we are in a better position to state the content of the three articles in chapter 2, 3 and 4. Each of them is directly or indirectly motivated by the problem of the semiclassical analysis of LQG.

1.6.1 Problem of time and general boundaries

Given that one has a proposal for transition amplitudes by means of spin foam sums, how would one encode the fact that initial and final states refer to certain physical times or have a certain physical time interval between them? An element of $\mathcal{H}_{\mathrm{kin}}$ or $\mathcal{H}_{\mathrm{diff}}$ contains information about quantum 3-geometries, but it does not specify the proper time that is associated to a transition process from one 3-geometry to another 3-geomery.

It has been proposed [7, 10] that this missing information could be encoded by using closed boundaries, in place of the 3d hypersurfaces of a foliation (see Fig. 1.8). On a closed boundary, the configuration consists of spatial 3-geometries g_{f}, g_{i}, and a timelike 3-geometry g_{t}. Along the time-like part, we can impose conditions on the proper distance between the space-like surfaces Σ_{i} and Σ_{f}. This distance corresponds to the proper time that is measured by clocks with world lines on Σ_{t}.

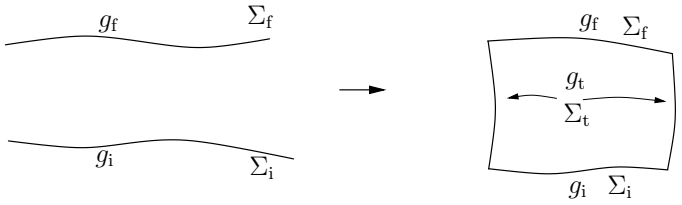


Figure 1.8: Transition from spacelike hypersurfaces to a closed boundary Σ.

Similarly, closed boundaries can be also introduced in standard field theory, where the geometry is fixed and non-dynamical: then, the choice of the boundary within the background determines the space-like and time-like distances on Σ.

In the research article of chapter 2, we investigate if a formulation with general boundaries can be implemented in the simplest possible context: that of a free scalar field theory in a Euclidean flat background. We arrive at a proposal for a discretized path integral for general boundaries, and an evolution equation—a generalized Schrödinger equation—that results from variation of the boundaries of the path integral.

Of course, the actual aim would be to use general boundaries in a background independent and Lorentzian context. The work of this article should be seen as a first check, if it is at all possible to make the general boundary formalism mathematically concrete.

In collaboration with others, we applied the general boundary idea also to spin foam gravity and made a formal proposal for a vacuum state [10]. From my present point of view, I consider it likely that this proposal is incorrect, so I do not include it in this thesis.

1.6.2 Semiclassical states for canonical LQG

The aim of a semiclassical analysis would be to show that, for suitable choices of initial and final states, the transition amplitudes of LQG reduce to the transition amplitudes of an effective field theory on a background geometry. If this effective field theory contained gravity, this would be the prove that LQG is indeed a quantum theory of gravity, and a candidate for the description of nature.

The big question is how the transition from fundamental to effective theory could be achieved. The fundamental theory is formulated in terms of spin network states. Evolution operators or spin foam sums define the transition amplitudes between these states. The degrees of freedom in such a formulation are labelled graphs and no longer fields, although they arise from a quantization of a field theory.

In the language of fields, it is relatively clear what we mean by semiclassicality: we mean that, for a suitable class of initial and final states, the dominant contributions come from configurations near a certain classical solution, and for that reason we expect there to be an effective field theory that has the classical configuration as a background and describes the transition amplitudes for our class of states. It is far from clear how this physical picture translates into a formalism where the basic configurations are spin networks and spin foams.

Thus, it is of central importance to understand how semiclassical properties manifest themselves at the level of the new degrees of freedom. We can view this problem as a result of a shift in language: by replacing fields with network-like degrees of freedom, we hope to resolve the UV divergences of field theory, but in this new language it is no longer evident how semiclassical approximations work.

The second research article of the thesis deals exactly with this translation problem. The strategy is to start from a vacuum state of linearized ADM gravity, which is formulated in

terms of fields, and translate it into state of canonical LQG. The translation procedure uses the fact that both network and field degrees of freedom are physically related. It is guided by the aim of preserving the framework-independent physical properties of the original ADM state.

The semiclassical peakedness of the ADM vacuum is mapped into certain properties of the LQG state, and this gives us an idea of how semiclassicality "looks like" in terms of spin networks. These properties provide us, in turn, with hints on how a semiclassical expansion around a background could be effected in canonical LQG.

1.6.3 Geometric spin foams and background-independence in spin foam models

We have seen that, in contrast to canonical LQG, the spin foam approach can already provide a definition of transition amplitudes. In this sense, spin foam gravity is closer to the point where a semiclassical analysis could be applied.

There exists, however, a problem in the definition of the spin foam sums: it usually involves a choice of triangulation or lattice κ, and that clashes with background-independence—the very thing that the theory promises to give!

In the third chapter of the thesis, we propose a solution to this difficulty. We introduce a symmetry condition which demands that the amplitude of an individual spin foam depends only on its geometric properties and not on the lattice on which it is defined. For models that have this property, we define a new sum over abstract spin foams that is independent of any choice of lattice or triangulation. We show furthermore that a version of the Barrett-Crane model satisfies this symmetry requirement.

As a side result, we obtain a simple dual formulation of lattice Yang-Mills theory that has a striking similarity with lattice quantizations of the Nambu-Goto string.

Chapter 2

Generalized Schrödinger equation in Euclidean field theory

2.1 Introduction

In quantum field theory (QFT) on Minkowski space, we can use the Schrödinger picture and have states associated to flat spacelike (hyper-)surfaces. The transition amplitude between an initial state and a final state is obtained by acting with the unitary evolution operator on the former and taking the inner product with the latter. An analogous Schrödinger picture has been also considered for QFT on curved spacetime [45]. In this case, states live on arbitrary spacelike Cauchy surfaces which form a foliation of spacetime. Evolution along these surfaces is non-unitary in general, as it does not correspond to a symmetry of the metric. When we come to a quantum theory of gravity, the spacetime geometry should become dynamical and there is no longer any fixed background metric; states live on arbitrary Cauchy surfaces and the requirement that the surface is spacelike is encoded in the state itself, which represents a quantum state of a spacelike geometry.

In all these cases, transition amplitudes are calculated for boundary states (i.e. an initial and final state) that are defined on spacelike boundaries. Recently, Oeckl has suggested that this restriction to *spacelike* boundaries could be relaxed [6, 7]. Oeckl offers heuristic arguments which suggest that transition amplitudes can be associated to a wider class of boundaries, as we do in topological quantum field theory [46]. These more general boundaries may include hypersurfaces which are partially timelike, that enclose a finite region of spacetime, or disjoint unions of such sets. This would imply, for instance, that in theories like QED or QCD, we could associate quantum "states" to a hypersphere, a hypercube or more exotic surfaces, and assign probability amplitudes to them. Similar suggestions were made in [11], with different motivations.

This "general boundary" approach to QFT could be interesting for several reasons. Firstly, finite closed boundaries represent the way real experiments are set up more directly than constant-time surfaces. A realistic experiment is confined to a finite region of spacetime. In particle colliders, for instance, the interaction region is enclosed by a finite outer region where state preparation and measurement take place. As sketched in Fig. 2.1, the walls and openings of a particle detector trace out a hypercube in spacetime. A "state" on the hypercube's surface would represent both incoming beams and jets of outgoing particles in a completely local fashion, without making any reference to inaccessible infinitely distant regions.

Secondly, in a quantum theory of gravity closed boundaries may provide a way to define scattering amplitudes, and help in solving the traditional interpretational difficulties of background independent QFT. This idea has been recently studied in [10], where it has also

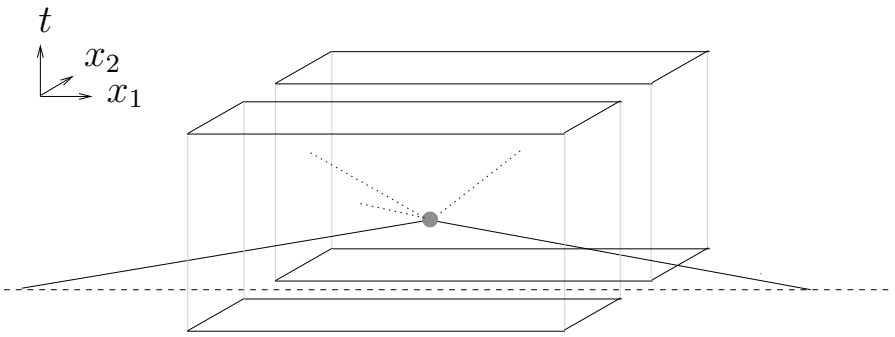


Figure 2.1: Spacetime diagram of particle scattering.

been used to propose an explicit way for computing the Minkowski vacuum state from a spin foam model. In a background independent theory the conventional spacelike states do not impose any constraint on the proper time elapsed between the initial and final states. As a result, the transition amplitude stems from a superposition of processes whose duration may range from microscopic to cosmic time scales. By means of timelike boundary conditions we could restrict this superposition to transitions of a certain duration and thereby implement the fact that experiments always involve clocks and proceed during a specified proper time interval.

Furthermore, in spin foam approaches, the introduction of general boundaries might open up the possibility of quantizing 3-geometries along time-like surfaces and clarify the physical meaning of Lorentzian spin foams.

Finally, a general boundary formulation could give us a broader perspective on QFT: it would stress geometrical aspects of QFT by no longer singling out a special subclass of surfaces, and may shed some light on the holographic principle, which states that the complete information about a spacetime region can be encoded in its boundary.

As noted by Oeckl, a heuristic idea for adapting QFT to general boundaries is provided by Feynman's sum-over-paths-picture. Given an arbitrary spacetime region V, bounded by a 3d hypersurface, the Feynman path integral over the spacetime region V, with fixed boundary value φ of the field, defines a functional $W[\varphi, V]$. This functional can be seen as a generalized evolution kernel, or a generalized field propagator. The path integral is therefore a natural starting point for developing a general boundary formalism.

The path to make these ideas precise is long. There are two types of problems. Firstly, the probabilistic interpretation of quantum theory and QFT must be adapted to this more general case. The physical meaning of states at fixed time and their relation to physical measurements are well established; the extension to arbitrary boundaries is probably doable, but far from obvious. It requires us to treat quantum state preparation and quantum measurement on the same ground, and to give a precise interpretation to the general probability amplitudes. Some steps in this direction can be found in [7] and [11].

Secondly, the mathematical apparatus of QFT, i.e. the path integral and operator formalism, needs to be extended to general spacetime regions. On a formal level, such a generalization appears natural for path integrals, but it is far from clear that it can be given a concrete and well-defined meaning.

In this paper we focus on the second of these issues: the definition of the field theoretical functional integral over an arbitrary region, and its relation to operator equations. We start to address the problem by considering the simplest system: Euclidean free scalar field theory. In this context, we propose an exact definition for the propagator kernel $W[\varphi, V]$,

based on limits of lattice path integrals. Under a number of assumptions, we can show that the propagator satisfies a generalized Schrödinger equation, of the Tomonaga-Schwinger kind [47, 48]. The equation governs the way the propagator changes under infinitesimal deformations of V. It reduces to the ordinary Schrödinger equation in the case in which a boundary surface of V is a constant-time surface and the deformation is a global shift in time.

With this result, we provide a first step towards constructing an operator formalism for general boundaries. The derivation can be seen as a higher-dimensional generalization of Feynman's path integral derivation of the Schrödinger equation for a single particle [49]. The main assumption we need is the existence of a rotationally invariant continuum limit.

We also derive the classical counterpart of the evolution equation: a generalized version of the Euclidean Hamilton-Jacobi equation. At present, we have no prescription for Wick rotation, so we cannot give any Lorentzian form for the propagator or the evolution equation. Hints in this direction were given in [10].

If one continues along this line, the ultimate goal would be to construct a full general boundary formalism for background dependent QFTs, which incorporates Wick rotation, interactions and renormalization. While of interest in itself, such a project could be also viewed as a testing ground for the general boundary method, which would prepare us for applying it in the more difficult context of background free QFT: there, as indicated before, the use of generalized boundary conditions may not only be helpful, but also essential for the physical analysis of the theory.

Our technique for deriving the evolution equation could be of interest in view of the attempts to relate canonical and path integral formulations of quantum gravity, i.e. when deriving the Wheeler-DeWitt equation from a concrete realization of a sum over geometries. (For existing results on this problem, see e.g. [50].)

The paper is organized as follows. In section 2.2, we present some of the heuristic considerations about state functionals on general boundaries, and their associated evolution kernel. Section 2.3 deals with the classical case: we present two derivations of the generalized Hamilton-Jacobi equation. The lattice regularization of the quantum propagator is defined in section 2.4. There, we also state the assumptions which are then used in section 2.5 for deriving the generalized Schrödinger equation. Both Hamilton-Jacobi and Schödinger equation are given in their integral form. In the appendix we clarify the relation with the local notation in [11].

Notation. V is the spacetime domain over which the action and the path integrals are defined. Σ is the boundary of V. The letter ϕ denotes a real scalar field on V, while φ stands for its restriction to Σ, i.e. $\varphi = \phi|_\Sigma$. Depending on the context, ϕ can be a solution of the classical equations of motion or an arbitrary field configuration. The action associated to ϕ is written as $S[\phi, V]$. When boundary conditions (φ,Σ) determine a classical solution ϕ on V, we denote the corresponding value of the action by $S[\varphi, V]$. Thus, the functional $S[\varphi, V]$ can be viewed as a Hamilton function (see sec. 3.3 of [11]). Vector components carry greek indices $\mu, \nu, \ldots$ (e.g. $v = (v^\mu)$). The dimension of spacetime is d. The symbol $\int_V \mathrm{d}^d x$ represents integrals over V, while integrals over Σ are indicated by $\int_\Sigma \mathrm{d}\Sigma(x)$. The letter n denotes the outward pointing and unit normal vector of Σ. The normal derivative is written as ∂_n, while ∇_Σ is the gradient along Σ. Accordingly, the full gradient ∇ decomposes on Σ as

$$\nabla|_\Sigma = n\,\partial_n + \nabla_\Sigma\,. \tag{2.1}$$

In section 2.4, we introduce a lattice with lattice spacing a and regularize various continuum quantities. Their discrete analogues are designated by the index a: for example, φ, V and Σ become φ_a, V_a and Σ_a.

2.2 General Boundary Approach

What is the meaning of a state on a general surface which is not necessarilly spacelike? What does it mean to propagate fields along a general spacetime domain? Following [7], an intuitive answer to these questions is provided by the path integral approach to QFT. We illustrate the intuitive idea in this section, as a heuristic motivation for the more rigorous definitions and developments in the remainder of the article. For simplicity, we refer here to a scalar field theory, but similar considerations can be extended to various path integral formulations of QFT, including sum over metrics or spin foam models in quantum gravity.

Consider Minkowskian scalar QFT in the Schrödinger picture. Let $|\Psi_{\rm i}\rangle$ be an initial state at time $t_{\rm i}$ and $|\Psi_{\rm f}\rangle$ a final state at time $t_{\rm f}$. The transition amplitude between the two is

$$A = \langle\Psi_{\rm f}|{\rm e}^{-{\rm i}H(t_{\rm f}-t_{\rm i})/\hbar}|\Psi_{\rm i}\rangle\,. \tag{2.2}$$

Using the functional representation, the amplitude (2.2) can be expressed as a convolution

$$A = \int D\varphi_{\rm f} \int D\varphi_{\rm i}\, \Psi_{\rm f}^*[\varphi_{\rm f}]\, W[\varphi_{\rm f}, t_{\rm f}; \varphi_{\rm i}, t_{\rm i}]\, \Psi_{\rm i}[\varphi_{\rm i}] \tag{2.3}$$

with the propagator kernel

$$W[\varphi_{\rm f}, t_{\rm f}; \varphi_{\rm i}, t_{\rm i}] := \langle\varphi_{\rm f}|{\rm e}^{-{\rm i}H(t_{\rm f}-t_{\rm i})/\hbar}|\varphi_{\rm i}\rangle\,. \tag{2.4}$$

This field propagator is a functional of the field: it should not be confused with the Feynman propagator, which is a two-point function, and propagates particles. When rewritten as a path integral, this kernel takes the form

$$W[\varphi_{\rm f}, t_{\rm f}; \varphi_{\rm i}, t_{\rm i}] = \int\limits_{\substack{\phi(.,t_{\rm i})=\varphi_{\rm i}\,,\\ \phi(.,t_{\rm f})=\varphi_{\rm f}}} D\phi\ {\rm e}^{{\rm i}S[\phi,t_{\rm i},t_{\rm f}]/\hbar}\,. \tag{2.5}$$

The action integral extends over the spacetime region $V_{\rm fi} := \mathbb{R}^{d-1} \times [t_{\rm i}, t_{\rm f}]$ and the path integral sums over all field configurations ϕ on $V_{\rm fi}$ that coincide with the fields $\varphi_{\rm f}$ and $\varphi_{\rm i}$ on the boundary. The complete boundary consists of two parts: the hyperplane $\Sigma_{\rm i}$ at the initial time $t_{\rm i}$, and the hyperplane $\Sigma_{\rm f}$ at the final time $t_{\rm f}$. We call their union $\Sigma_{\rm fi} := \Sigma_{\rm f} \cup \Sigma_{\rm i}$. If we view $\varphi_{\rm f}$ and $\varphi_{\rm i}$ as components of a single boundary field $\varphi_{\rm fi} := (\varphi_{\rm f}, \varphi_{\rm i})$ on $\Sigma_{\rm fi}$, we can write the evolution kernel (2.5) more concisely as

$$W[\varphi_{\rm fi}, V_{\rm fi}] := \int_{\phi|_{\Sigma_{\rm fi}}=\varphi_{\rm fi}} D\phi\ {\rm e}^{{\rm i}S[\phi,V_{\rm fi}]/\hbar}\,.$$

With this notation, it seems natural to introduce a propagator functional for more general spacetime regions V (see Fig. 2.2): we define it as

$$W[\varphi, V] := \int_{\phi|_{\Sigma}=\varphi} D\phi\ {\rm e}^{{\rm i}S[\phi,V]/\hbar}\,. \tag{2.6}$$

Here ϕ varies freely on the interior of V and is fixed to the value φ on the boundary Σ. Of course, this is only a formal expression, and it is not clear that it can be given mathematical meaning. Let us suppose for the moment that it *has* meaning and see what would follow from it.

Ordinary propagators satisfy convolution (or Markov) identities which result from the subdivision or joining of time intervals. If the functional W behaves the way our naive picture

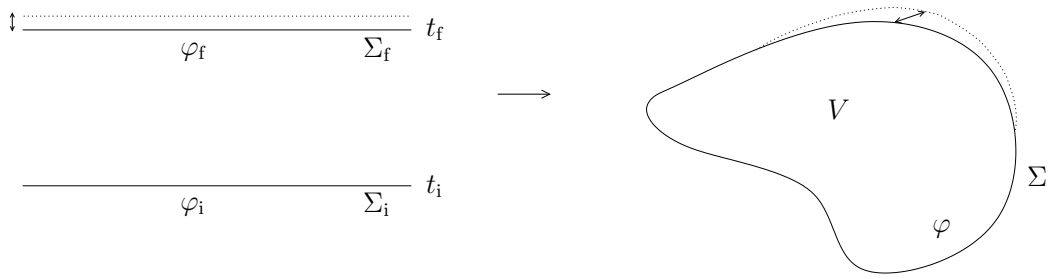


Figure 2.2: From $V_{\rm fi}$ to general V.

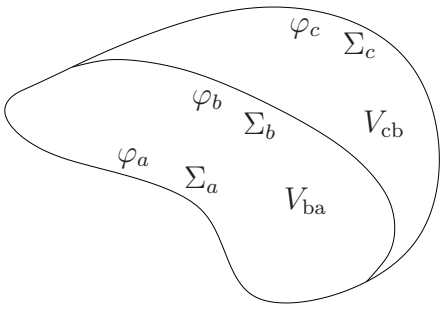


Figure 2.3: Splitting of V.

tells us, the splitting and joining of volumes should translate into analogous convolution relations. For instance, if V is divided as shown Fig. 2.3, the new regions $V_{\rm cb}$ and $V_{\rm ba}$ carry propagators

$$W[(\varphi_{\rm c},\varphi_{\rm b}),V_{\rm cb}] \;=\; \int\limits_{\phi|_{\Sigma_{\rm cb}}=(\varphi_{\rm c},\varphi_{\rm b})} D\phi\; {\rm e}^{{\rm i}S[\phi,V_{\rm cb}]/\hbar}\,, \tag{2.7}$$

$$W[(\varphi_{\rm b},\varphi_{\rm a}),V_{\rm ba}] \;=\; \int\limits_{\phi|_{\Sigma_{\rm ba}}=(\varphi_{\rm b},\varphi_{\rm a})} D\phi\; {\rm e}^{{\rm i}S[\phi,V_{\rm c}]/\hbar}\,. \tag{2.8}$$

When integrating the product of (2.7) and (2.8) over the field $\varphi_{\rm b}$ along the common boundary, one recovers the original propagator:

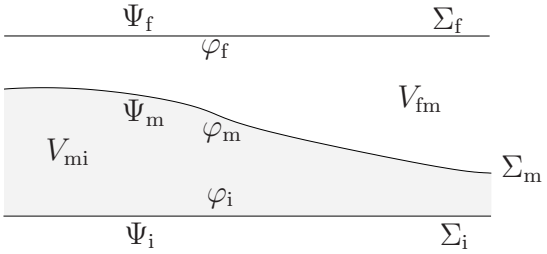


Figure 2.4: Splitting of $V_{\rm fi}$.

$$W[(\varphi_{\rm c},\varphi_{\rm a}),V] = \int D\varphi_{\rm b}\, W[(\varphi_{\rm c},\varphi_{\rm b}),V_{\rm cb}]\, W[(\varphi_{\rm b},\varphi_{\rm a}),V_{\rm ba}]\,. \tag{2.9}$$

Similarly, the infinite strip $V_{\rm fi}$ between $t_{\rm f}$ and $t_{\rm i}$ could be cut by a "middle" surface $\Sigma_{\rm m}$ as in Fig. 2.4, giving the new volumes $V_{\rm fm}$ and $V_{\rm mi}$. Its kernel decomposes as

$$W[(\varphi_{\rm f},\varphi_{\rm i}),V_{\rm fi}] = \int D\varphi_{\rm m}\, W[(\varphi_{\rm f},\varphi_{\rm m}),V_{\rm fm}]\, W[(\varphi_{\rm m},\varphi_{\rm i}),V_{\rm mi}]\,.$$

Thus, the evolution of $|\Psi_{\rm i}\rangle$ to the final time is divided into two steps: using the propagator on $V_{\rm mi}$, we evolve up to the surface $\Sigma_{\rm m}$ and obtain the intermediate state

$$\Psi_{\rm m}[\varphi_{\rm m}] := \int {\rm d}\varphi_{\rm i}\, W[(\varphi_{\rm m},\varphi_{\rm i}),V_{\rm mi}]\, \Psi_{\rm i}[\varphi_{\rm i}]\,.$$

The kernel $W[.,V_{\rm fm}]$ covers the remaining evolution and gives the original amplitude (2.2) when convoluted with $\Psi_{\rm f}^*$ and $\Psi_{\rm m}$

$$A = \int D\varphi_{\rm f} \int D\varphi_{\rm m}\, \Psi_{\rm f}^*[\varphi_{\rm f}]\, W[(\varphi_{\rm f},\varphi_{\rm m}),V_{\rm fm}]\, \Psi_{\rm m}[\varphi_{\rm m}]\,. \tag{2.10}$$

Since the same amplitude can either be calculated from $\Psi_{\rm f}$ and $\Psi_{\rm i}$ or from $\Psi_{\rm f}$ and $\Psi_{\rm m}$, the wave functional $\Psi_{\rm m}$ encodes all physical information about the intial state. On account of this property, we say that $\Psi_{\rm m}$ *is the state functional which results from evolving* $\Psi_{\rm i}$ *by the volume* $V_{\rm mi}$. As for ordinary state functionals, one can think of $\Psi_{\rm m}$ as being an element $|\Psi_{\rm m}\rangle$ in a Hilbert space, which we call $\mathcal{H}_{\Sigma_{\rm m}}$. The latter consists of functionals of fields over $\Sigma_{\rm m}$ and has the inner product

$$\langle\Psi_2|\Psi_1\rangle := \int D\varphi_{\rm m}\, \Psi_2^*[\varphi_{\rm m}]\Psi_1[\varphi_{\rm m}]\,, \qquad |\Psi_1\rangle\,,\, |\Psi_2\rangle \in \mathcal{H}_{\Sigma_{\rm m}}\,.$$

It is important to note that the evolution map from $\mathcal{H}_{\Sigma_{\rm i}}$ to $\mathcal{H}_{\Sigma_{\rm m}}$ need not be unitary. The results of Torre and Varadarajan show, in fact, that in flat spacetime state evolution between curved Cauchy surfaces cannot be implemented unitarily [51]. Nevertheless, a probability interpretation *is* viable for states in $H_{\Sigma_{\rm m}}$, as the meaning of amplitudes such as (2.10) can be traced back to that of the standard amplitude (2.3).

Consider now a more unconventional example. Cut out a bounded and simply connected set $V_{\rm m}$ from $V_{\rm fi}$ and denote the remaining volume by $V_{\rm fmi}$ (Fig. 2.5). This time we define the state functional $\Psi_{\rm m}$ by "evolving" both $\Psi_{\rm f}$ and $\Psi_{\rm m}$ to the middle boundary $\Sigma_{\rm m}$, i.e.

$$\Psi_{\rm m}[\varphi_{\rm m}] := \int D\varphi_{\rm f} \int D\varphi_{\rm i}\, \Psi_{\rm f}^*[\varphi_{\rm f}]\, W[(\varphi_{\rm f},\varphi_{\rm m},\varphi_{\rm i}),V_{\rm fmi}]\, \Psi_{\rm i}[\varphi_{\rm i}]\,.$$

Clearly, the amplitude (2.3) is now equal to

$$A = \int D\varphi_{\rm m}\, W[\varphi_{\rm m},V_{\rm m}]\, \Psi_{\rm m}[\varphi_{\rm m}]\,. \tag{2.11}$$

Therefore, the functional $\Psi_{\rm m}$ contains the entire information needed to compute the transition amplitude between $\Psi_{\rm i}$ and $\Psi_{\rm f}$.

To make this more concrete and more intuitive, suppose that the scalar field theory is free and that $\Psi_{\rm i}$ and $\Psi_{\rm f}$ are the initial and final one-particle states of a single, localized particle whose (smeared out) worldline passes through $V_{\rm m}$. In both functionals, the presence of the particle appears as a local deviation from the vacuum, in the functional dependence. Likewise, it is natural to presume that the functional form of $\Psi_{\rm m}$ reflects where the worldline of the particle enters and exits the volume $V_{\rm m}$.

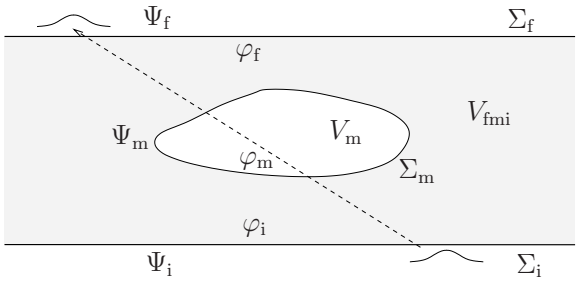


Figure 2.5: Evolution to a closed surface Σ_m.

How can we interpret the "state" Ψ_m and the associated amplitude (2.11)? To answer this, let us get back to equation (2.2). Notice that the amplitude A depends on the *couple* of states $(|\Psi_\mathrm{i}\rangle, |\Psi_\mathrm{f}\rangle)$. This couple represents a possible outcome of a measurement at time t_f as well as a state preparation at time t_i. A state preparation is itself a quantum measurement, therefore we can say that this couple represents a possible outcome of an ensemble of quantum measurements performed at times t_i and t_f. We may introduce a name to denote such a couple. We call it a *process*, since the two states $(|\Psi_\mathrm{i}\rangle, |\Psi_\mathrm{f}\rangle)$, taken together, represent the ensemble of data (initial and final) that we can gather about a physical process. A probability amplitude is associated to the entire process $(|\Psi_\mathrm{i}\rangle, |\Psi_\mathrm{f}\rangle)$. Now, it is clear that the functional Ψ_m represents a generalization of this idea of a process. It is tempting to presume that Ψ_m can be interpreted as representing a possible outcome of quantum measurements that can be made on Σ_m. In the example of the particle above, for instance, it will represent the detection of the incoming and outgoing particle.

The idea is that given an arbitrary closed surface, the possible results of the ensemble of measurements that we can make on it determines a space of generalized "states" which can be associated to the surface. Each such state represents a process whose probabilistic amplitude is provided by expression (2.11). The conventional formalism is recovered when the surface is formed by two parallel spacelike planes. For more details on the physical interpretation of general boundary states, see sec. 5.3 of [11].

2.2.1 Operator Formalism

If path integrals can be defined for general boundaries, how would a corresponding operator formalism look like? In particular, is there an operator that governs the dynamics, as the Hamiltonian does for rigid time translations? Recall that the Hamiltonian can be recovered from the path integral by considering an infinitesimal shift of the final time. For example, if we displace by a time interval Δt the final surface Σ_f in (2.5), keeping the same boundary field φ_f, the new propagator results from the convolution

$$W[\varphi_\mathrm{f}, t_\mathrm{f} + \Delta t; \varphi_\mathrm{i}, t_\mathrm{i}] = \int D\varphi\, W[\varphi_\mathrm{f}, t_\mathrm{f} + \Delta t; \varphi, t_\mathrm{f}]\, W[\varphi, t_\mathrm{f}; \varphi_\mathrm{i}, t_\mathrm{i}]\,. \tag{2.12}$$

For infinitesimal Δt, this gives the Schrödinger equation, which expresses the variation of W in terms of the Hamiltonian operator

$$\left(\mathrm{i}\hbar\frac{\partial}{\partial t_\mathrm{f}} - H[\varphi_\mathrm{f}, -\mathrm{i}\hbar\frac{\delta}{\delta\varphi_\mathrm{f}}]\right) W[\varphi_\mathrm{f}, t_\mathrm{f}; \varphi_\mathrm{i}, t_\mathrm{i}] = 0\,, \tag{2.13}$$

where

$$H[\varphi_\mathrm{f}, \pi_\mathrm{f}] \;=\; \int_{\Sigma_\mathrm{f}} \mathrm{d}\Sigma\; \frac{1}{2}\left(\pi_\mathrm{f}^2 + (\nabla_\Sigma\varphi_\mathrm{f})^2 + m^2\varphi_\mathrm{f}^2\right)\,. \tag{2.14}$$

Similarly, if φ_{f} is displaced in a tangential direction $e_{\|}$ along Σ_{f}, the variation of W is generated by the momentum operator

$$e_{\|} \cdot P[\varphi_{\mathrm{f}}, -\mathrm{i}\hbar \frac{\delta}{\delta \varphi_{\mathrm{f}}}] \,,$$

where

$$P[\varphi_{\mathrm{f}}, \pi_{\mathrm{f}}] = -\int_{\Sigma_{\mathrm{f}}} \mathrm{d}\Sigma(x)\, \nabla_{\Sigma} \varphi_{\mathrm{f}}(x)\, \pi_{\mathrm{f}} \,.$$

In the case of a general volume V, it is natural to expect that deformations of the boundary surface Σ lead to an analogous functional differential equation for the propagator. However, for a general shape of V there is no notion of preferred rigid displacement of the boundary. We must consider arbitrary deformations of the boundary surface, and we expect that the associated change in W is governed by a generalized Schrödinger equation (see Fig. 2.2). In the same way that H and P generate temporal and spatial shifts, the operators in such a Schrödinger equation could be seen as the generators for general boundary deformations of W. In a diffeomorphism invariant QFT, the analogous W-functional would be independent of Σ, and the generalized Schrödinger equation reduces to the Wheeler-DeWitt equation.

In the present paper, we consider only Euclidean field theory, so we seek to define the Euclidean form

$$W[\varphi, V] := \int_{\phi|_{\Sigma}=\varphi} D\phi \; \mathrm{e}^{-S[\phi,V]/\hbar} \,. \tag{2.15}$$

of the propagator (2.6), and generalize the Euclidean version of the Schrödinger equation (2.13).

Before dealing with path integrals and deformations of their boundaries, however, we discuss the analogous problem in classical field theory. The classical counterpart of the Schrödinger equation is the Hamilton-Jacobi equation. The Hamilton function $S[\varphi_{\mathrm{f}}, t_{\mathrm{f}}, \varphi_{\mathrm{i}}, t_{\mathrm{i}}]$ is a function of the same arguments as the field propagator (2.4). It is defined as the value of the action of the classical field configuration which solves the equations of motion and has the given boundary values. It satisfies the Hamilton-Jacobi equation

$$\frac{\partial}{\partial t_{\mathrm{f}}} S[\varphi_{\mathrm{f}}, t_{\mathrm{f}}; \varphi_{\mathrm{i}}, t_{\mathrm{i}}] + H[\varphi_{\mathrm{f}}, \frac{\delta S}{\delta \varphi_{\mathrm{f}}}] = 0. \tag{2.16}$$

For more general regions V, the Hamilton function becomes a functional of V and the boundary field φ specified on Σ. In the next section we show that this functional satisfies a generalized Hamilton-Jacobi equation which governs its dependence on arbitrary variations of V.

2.3 Generalized Hamilton-Jacobi Equation

Let V be an open and simply connected subset of Euclidean d-dimensional space $\mathbb{R}^d$. We consider the Euclidean action

$$S[\phi, V] = \int_V d^d x \left[\frac{1}{2} (\nabla \phi)^2 + \frac{1}{2} m^2 \phi^2 + U(\phi) \right] , \tag{2.17}$$

where U is some polynomial potential in ϕ. In the classical case, unlike in the quantum case, an interaction term can be added without complicating the derivation that follows. The equations of motion are

$$\Box \phi - m^2 \phi - \frac{\partial U}{\partial \phi} = 0 \,. \tag{2.18}$$

The Hamilton function $S[\varphi, V]$ is defined by $S[\varphi, V] = S[\phi, V]$, where ϕ solves (2.18) and $\phi|_\Sigma = \varphi$. It is defined for all values (φ, Σ) where this solution exists and is multivalued if this solution is not unique.

We now study the change in $S[\varphi, V]$ under a local variation of V. To make this precise, consider a vector field $N = (N^\mu)$ over $\mathbb{R}^d$. N induces a flow on $\mathbb{R}^d$ which we denote by $\sigma : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$. Define the transformed volume as $V^s := \sigma(s, V)$. Likewise, $\Sigma^s := \sigma(s, \Sigma)$.

To define the change in $S[\varphi, V]$ under a variation of V, we need to specify what value the boundary field should take on the new boundary Σ^s. We choose it to be the pull-forward by $\sigma_s \equiv \sigma(s, .)$, i.e. $\varphi^s := \varphi \circ \sigma_s^{-1} \equiv \sigma_{s*}\varphi$. Let us assume that the point (φ, Σ) is regular in the space of boundary conditions, in the sense that slightly deformed boundary conditions (φ^s, Σ^s) give a new unique solution ϕ^s on V^s, close to the previous one. In this case, the number $S[\varphi^s, V^s]$ is well-defined and we can write down the differential quotient

$$L_N S[\varphi, V] := \lim_{s \to 0} \frac{1}{s} \left(S[\varphi^s, V^s] - S[\varphi, V] \right) , \tag{2.19}$$

with the vector field N as a parameter. As we show below, this limit exists and the map L_N is a functional differential operator. The local form of this differential operator is given in the appendix.

We decompose the restriction of N to Σ into its components normal and tangential to Σ,

$$N_{|\Sigma} = N_\perp n + N_\parallel ,$$

where the scalar field $N_\perp$ is defined as

$$N_\perp := n_\mu N^\mu .$$

Observe that under a small variation $\delta\varphi$ of the boundary field, we have

$$\begin{aligned}
\delta S[\varphi, V] &= S[\delta\varphi, V] = S[\delta\phi, V] \\
&= \int_V d^d x \left[\nabla^\mu \delta\phi \, \nabla_\mu \phi + m^2 \phi \, \delta\phi + \frac{\partial U}{\partial \phi} \delta\phi \right] \\
&= \int_V d^d x \Big[\nabla_\mu (\nabla_\mu \phi \, \delta\phi) + \delta\phi \underbrace{\left(-\Box\phi + m^2 \phi + \frac{\partial U}{\partial \phi} \right)}_{=0} \Big] \\
&= \int_\Sigma \mathrm{d}\Sigma \, \partial_n \phi \, \delta\phi = \int_\Sigma \mathrm{d}\Sigma \; \partial_n \phi \, \delta\varphi .
\end{aligned}$$

Therefore, we have

$$\frac{\delta S}{\delta \varphi(x)} [\varphi, V] = \partial_n \phi(x) . \tag{2.20}$$

2.3.1 Direct Derivation

Suppose for the moment that V is only extended by the deformation (i.e. $V \subset V^s$ for every s). Then, the most direct derivation of the Hamilton-Jacobi equation can be obtained by considering the restriction φ^s_Σ of ϕ^s to Σ: that is, the value of the classical solution on Σ when the boundary condition φ^s is specified on Σ^s (see Fig. 2.6). Note that $\varphi^0_\Sigma = \varphi^0 = \varphi$. By inserting $S[\varphi^s_\Sigma, V]$ into the difference, i.e.

$$S[\varphi^s, V^s] - S[\varphi, V] = S[\varphi^s, V^s] - S[\varphi^s_\Sigma, V] + S[\varphi^s_\Sigma, V] - S[\varphi, V] ,$$

the differential quotient becomes a sum of two limits:

$$\lim_{s \to 0} \frac{S[\varphi^s, V^s] - S[\varphi, V]}{s} = \lim_{s \to 0} \frac{S[\varphi^s, V^s] - S[\varphi^s_\Sigma, V]}{s} + \lim_{s \to 0} \frac{S[\varphi^s_\Sigma, V] - S[\varphi, V]}{s}$$

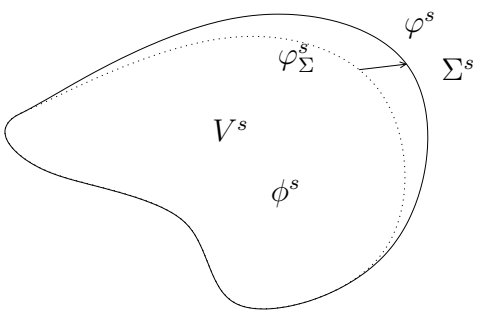


Figure 2.6: Definition of φ^s_Σ.

As φ^s and φ^s_Σ are part of the same solution, the first limit is easily seen to be

$$\begin{aligned}\lim_{s\to 0}\frac{S[\varphi^s,V^s]-S[\varphi^s_\Sigma,V]}{s} &= \int_\Sigma \mathrm{d}\Sigma\; N_\perp\left(\frac{1}{2}\nabla^\mu\phi\nabla_\mu\phi+\frac{1}{2}m^2\phi^2+U(\phi)\right)\\ &= \int_\Sigma \mathrm{d}\Sigma\; N_\perp\left(\frac{1}{2}\left(\frac{\delta S}{\delta\varphi}\right)^2+\frac{1}{2}(\nabla_\Sigma\phi)^2+\frac{1}{2}m^2\phi^2+U(\phi)\right)\,.\end{aligned}$$

In the last line, we used the decomposition (2.1) and equation (2.20). The second differential quotient gives

$$\begin{aligned}\lim_{s\to 0}\frac{S[\varphi^s_\Sigma,V]-S[\varphi,V]}{s} &= \int_\Sigma \mathrm{d}\Sigma\;\frac{\delta S}{\delta\varphi}\,\frac{\mathrm{d}}{\mathrm{d}s}\varphi^s_\Sigma|_{s=0}\\ &= \int_\Sigma \mathrm{d}\Sigma\;\frac{\delta S}{\delta\varphi}\left(-N_\perp\partial_n\phi-N_\parallel\cdot\nabla_\Sigma\varphi\right)\,.\end{aligned}$$

Altogether one has

$$L_N S[\varphi,V]=\int_\Sigma \mathrm{d}\Sigma\left\{N_\perp\left[-\frac{1}{2}\left(\frac{\delta S}{\delta\varphi}\right)^2+\frac{1}{2}(\nabla_\Sigma\phi)^2+\frac{1}{2}m^2\phi^2+U(\phi)\right]-N_\parallel\cdot\nabla_\Sigma\varphi\frac{\delta S}{\delta\varphi}\right\}\,. \tag{2.21}$$

We started from the assumption that $V\subset V^s$ for all s, but it is easy to see that the previous argument can be adapted to the general case where the volume V is partly extended and partly decreased.

If we introduce the quantities

$$\begin{aligned}H_N[\varphi,\pi,V] &:= \int_\Sigma \mathrm{d}\Sigma\; N_\perp\left(-\frac{1}{2}\pi^2+\frac{1}{2}(\nabla_\Sigma\varphi)^2+\frac{1}{2}m^2\varphi^2+U(\varphi)\right)\,,\\ P_N[\varphi,\pi,V] &:= -\int_\Sigma \mathrm{d}\Sigma\, N_\parallel\cdot\nabla_\Sigma\varphi\,\pi\,,\end{aligned}$$

equation (2.21) takes the form

$$L_N S[\varphi,V]=H_N[\varphi,\frac{\delta S}{\delta\varphi},V]+P_N[\varphi,\frac{\delta S}{\delta\varphi},V]\,. \tag{2.22}$$

When V is a strip of spacetime between times t_i and t_f, and $N_\perp|_{t_\mathrm{f}}=1$, $N_\perp|_{t_\mathrm{i}}=0$, $N_\parallel=0$, equation (2.22) reduces to the usual Hamilton-Jacobi equation

$$\frac{\partial}{\partial t_\mathrm{f}}S[\varphi_\mathrm{f},t_\mathrm{f};\varphi_\mathrm{i},t_\mathrm{i}]=\int_{\Sigma_\mathrm{f}}\mathrm{d}\Sigma\;\left(-\frac{1}{2}\left(\frac{\delta S}{\delta\varphi_\mathrm{f}}\right)^2+\frac{1}{2}(\nabla_\Sigma\varphi_\mathrm{f})^2+\frac{1}{2}m^2\varphi_\mathrm{f}^2+U(\varphi_\mathrm{f})\right)$$

of Euclidean field theory. Hence (2.22) can be seen as a geometric generalization of the Hamilton-Jacobi equation.

2.3.2 Alternative Derivation

Let us describe another way of evaluating the "deformation derivative" $L_N S[\varphi, V]$. The spacetime metric tensor g enters in the definition of the action and therefore in the definition of $S[\varphi, V]$. Let us write this dependence explicitly as $S[\varphi, g, V]$. A diffeomorphism that acts on ϕ, the boundary Σ *and* the metric g, leaves the action invariant. Therefore

$$S[\varphi^s, g^s, V^s] = S[\varphi, g, V]\,.$$

Equivalently,

$$S[\varphi^s, g, V^s] = S[\varphi, g^{-s}, V].$$

Plugging this into the definition of the operator (2.19) gives

$$L_N S[\varphi, g, V] = \lim_{s\to 0} \frac{1}{s}\left(S[\varphi, g^{-s}, V] - S[\varphi, g, V]\right)\,, \tag{2.23}$$

which is a variation of the action w.r.t. the metric *only*. Now we can use the definition of the energy-momentum tensor to obtain

$$\begin{aligned} L_N S[\varphi, g, V] &= \frac{1}{2}\int_V d^d x\, T^{\mu\nu} \frac{\mathrm{d}}{\mathrm{d}s} g^{-s}_{\mu\nu}|_{s=0} = \frac{1}{2}\int_V d^d x\, T^{\mu\nu}(-\nabla_\mu N_\nu - \nabla_\nu N_\mu) \\ &= -\int_V d^d x\, T^{\mu\nu}\nabla_\mu N_\nu = -\int_\Sigma \mathrm{d}\Sigma\, n_\mu T^{\mu\nu} N_\nu + \int_V \mathrm{d}^d x \underbrace{\nabla_\mu T^{\mu\nu}}_{=0} N_\nu \\ &= -\int_\Sigma \mathrm{d}^d x\, n_\mu T^{\mu\nu} N_\nu \end{aligned} \tag{2.24}$$

In the last two steps we used Stoke's theorem and the equations of motion respectively. On the other hand, we know that

$$T^{\mu\nu} = -g^{\mu\nu}\mathcal{L} + \nabla^\mu\phi\nabla^\nu\phi$$

and

$$\begin{aligned} n_\mu N_\nu T^{\mu\nu} &= -N_\perp\left[\frac{1}{2}(\partial_n\phi)^2 + \frac{1}{2}(\nabla_\Sigma\phi)^2 + \frac{1}{2}m^2\phi^2 + U(\phi)\right] + \partial_n\phi(N_\perp\partial_n\phi + N_\parallel\cdot\nabla_\Sigma\phi) \\ &= -N_\perp\left[-\frac{1}{2}(\partial_n\phi)^2 + \frac{1}{2}(\nabla_\Sigma\phi)^2 + \frac{1}{2}m^2\phi^2 + U(\phi)\right] + N_\parallel\cdot\nabla_\Sigma\phi\,\partial_n\phi \end{aligned} \tag{2.25}$$

Inserting (2.25) in (2.24) and using (2.20), we arrive again at the generalized Hamilton-Jacobi equation

$$L_N S[\varphi, V] = \int_\Sigma \mathrm{d}\Sigma\left\{N_\perp\left[-\frac{1}{2}\left(\frac{\delta S}{\delta\varphi}\right)^2 + \frac{1}{2}(\nabla_\Sigma\phi)^2 + \frac{1}{2}m^2\phi^2 + U(\phi)\right] - N_\parallel\cdot\nabla_\Sigma\varphi\frac{\delta S}{\delta\varphi}\right\}\,.$$

2.4 Definition of the Evolution Kernel

In this section, we define a Euclidean free field propagator for arbitrary spacetime domains V. Limits of lattice path integrals are used to give a precise meaning to the expression (2.15). We begin by considering the case $V = V_{\text{fi}}$ and derive the lattice path integral from the operator formalism. Then, we propose a way to extend this expression to more general volumes V.

2.4.1 From Operators to Path Integrals

The transition from operator formalism to path integral is a standard procedure. We repeat it here, since treatments of lattice field theory usually omit normalization factors. There, constant factors drop out when dividing by the partition function Z. In our case, their precise form will be crucial for the definition of the propagator.

In the Schrödinger picture, the space of states $\mathcal{H}$ is associated to the manifold $\mathbb{R}^{d-1}$: we regularize it by a finite lattice

$$S_a := \{\underline{x} \in a\mathbb{Z}^{d-1} \mid -Ma \leq |x_i| \leq Ma\,,\ i = 1, \ldots, d-1\}$$

with lattice constant $a > 0$ and edge length $2aM$, $M \in \mathbb{N}$. e_i is the unit vector in the ith direction. For a scalar function f on S_a, the forward derivative is

$$\nabla_i f(\underline{x}) := \frac{\phi(\underline{x} + ae_i) - \phi(\underline{x})}{a}\,,$$

and we set $\phi(\underline{x} + ae_i) := \phi(\underline{x} - aMe_i)$ when $x_i = aM$. Let $\{\hat{\phi}(\underline{x})\}$, $\{\hat{\pi}(\underline{x})\}$ be canonical operators with eigenstates $\{|\phi\rangle\}$, $\{|\pi\rangle\}$ such that

$$\hat{\phi}(\underline{x})\,|\,\phi\rangle = \phi(\underline{x})|\phi\rangle, \qquad \hat{\pi}(\underline{x})|\pi\rangle = \pi(\underline{x})|\pi\rangle\,, \qquad \underline{x} \in S_a\,, \tag{2.26}$$

and

$$[\hat{\phi}(\underline{x}), \hat{\pi}(\underline{y})] = \frac{\mathrm{i}\hbar}{a^{d-1}}\delta(\underline{x} - \underline{y})\,, \qquad \underline{x}, \underline{y} \in S_a\,. \tag{2.27}$$

The eigenstates are normalized as

$$\langle\phi, \phi'\rangle = \prod_{\underline{x} \in S_a} \delta(\phi(\underline{x}) - \phi'(\underline{x}))\,, \qquad \langle\pi, \pi'\rangle = \prod_{\underline{x} \in S_a} \delta(\pi(\underline{x}) - \pi'(\underline{x}))\,, \tag{2.28}$$

and give rise to completeness relations

$$\left(\prod_{\underline{x} \in S_a} \int_{-\infty}^{\infty} \mathrm{d}\phi(\underline{x})\right) |\phi\rangle\langle\phi| = \mathbb{1}\,, \qquad \left(\prod_{\underline{x} \in S_a} \int_{-\infty}^{\infty} \mathrm{d}\pi(\underline{x})\right) |\pi\rangle\langle\pi| = \mathbb{1}\,. \tag{2.29}$$

From (2.26), (2.27) and (2.28), it follows that

$$\hat{\pi}(\underline{x}) = -\frac{\mathrm{i}\hbar}{a^{d-1}} \frac{\partial}{\partial\phi(\underline{x})} \tag{2.30}$$

and

$$\langle\phi, \pi\rangle = \left(\prod_{\underline{x} \in S_a} \sqrt{\frac{a^{d-1}}{2\pi\hbar}}\right) \exp\left(\frac{\mathrm{i}}{\hbar} \sum_{\underline{x} \in S_a} a^{d-1}\phi(\underline{x})\pi(\underline{x})\right) \tag{2.31}$$

The Hamiltonian operator is

$$\begin{aligned} H[\hat{\phi}, \hat{\pi}] &:= \sum_{\underline{x} \in S_a} a^{d-1} \left[\frac{1}{2}\hat{\pi}^2(\underline{x}) + \frac{1}{2}(\underline{\nabla}\hat{\phi})^2(\underline{x}) + \frac{1}{2}m^2\hat{\phi}^2(\underline{x})\right] \\ &\equiv T[\hat{\pi}] + V[\hat{\phi}]\,. \end{aligned}$$

We rewrite the Euclidean propagator

$$\langle\varphi_{\mathrm{f}}|\mathrm{e}^{-H(t_{\mathrm{f}} - t_{\mathrm{i}})/\hbar}|\varphi_{\mathrm{i}}\rangle\,, \qquad t_{\mathrm{f}} - t_{\mathrm{i}} = na\,,$$

by inserting repeatedly the completeness relations (2.29):

$$\langle\varphi_{\rm f}|{\rm e}^{-naH/\hbar}|\varphi_{\rm i}\rangle = \left(\prod_{k=1}^{n-1}\prod_{\underline{x}}\int {\rm d}\phi_k(\underline{x})\right)\left(\prod_{k=0}^{n-1}\prod_{\underline{x}}\int {\rm d}\pi_k(\underline{x})\right)\times$$
$$\times\,\langle\varphi_{\rm f}|\pi_{n-1}\rangle\langle\pi_{n-1}|{\rm e}^{-aH/\hbar}|\phi_{n-1}\rangle\langle\phi_{n-1}|\pi_{n-2}\rangle\langle\pi_{n-2}|{\rm e}^{-aH/\hbar}|\phi_{n-2}\rangle\cdots\langle\phi_1|\pi_0\rangle\langle\pi_0|{\rm e}^{-aH/\hbar}|\varphi_{\rm i}\rangle$$

After making the replacement

$${\rm e}^{-aH/\hbar} = {\rm e}^{-aT/\hbar}{\rm e}^{-aV/\hbar} + O(a^2) \;\to\; {\rm e}^{-aT/\hbar}{\rm e}^{-aV/\hbar}$$

and using (2.31), we obtain

$$\left(\prod_{k=1}^{n-1}\prod_{\underline{x}}\int {\rm d}\phi_k(\underline{x})\right)\left(\prod_{k=0}^{n-1}\prod_{\underline{x}}\int {\rm d}\pi_k(\underline{x})\right)\langle\varphi_{\rm f}|\pi_{n-1}\rangle\langle\pi_{n-1}|\phi_{n-1}\rangle\langle\phi_{n-1}|\pi_{n-2}\rangle\langle\pi_{n-2}|\phi_{n-2}\rangle\cdots$$
$$\cdots\langle\phi_1|\pi_0\rangle\langle\pi_0|\varphi_{\rm i}\rangle\exp\left(\frac{1}{\hbar}\sum_{k=0}^{n-1}aH[\phi_k,\pi_k]\right)\Bigg|_{\phi_0\,=\,\varphi_{\rm i}}$$
$$= \left(\prod_{k=1}^{n-1}\prod_{\underline{x}}\int {\rm d}\phi_k(\underline{x})\right)\left(\prod_{k=0}^{n-1}\prod_{\underline{x}}\int {\rm d}\pi_k(\underline{x})\right)\left(\prod_{\underline{x}}\sqrt{\frac{a^{d-1}}{2\pi\hbar}}\right)^{2n}\times$$
$$\times\,\exp\left\{\frac{1}{\hbar}\sum_{k=0}^{n-1}a\sum_{\underline{x}}a^{d-1}\left[{\rm i}\frac{\phi_{k+1}(\underline{x})-\phi_k(\underline{x})}{a}\pi_k(\underline{x})\right.\right.$$
$$\left.\left.-\;\frac{1}{2}\pi_k^2(\underline{x})-\frac{1}{2}(\underline{\nabla}\phi_k)^2(\underline{x})-\frac{1}{2}m^2\phi_k^2(\underline{x})\right]\right\}\Bigg|_{\substack{\phi_0\,=\,\varphi_{\rm i}\,,\\ \phi_n\,=\,\varphi_{\rm f}}}.$$

Integration over the momenta yields the path integral

$$\left(\prod_{k=1}^{n-1}\prod_{\underline{x}}\int {\rm d}\phi_k(\underline{x})\right)\left(\prod_{\underline{x}}\left(\frac{a^{d-1}}{2\pi\hbar}\right)^n\left(\frac{2\pi\hbar}{a^d}\right)^{n/2}\right)\times$$
$$\times\,\exp\left\{-\frac{1}{\hbar}\sum_{k=0}^{n-1}a\sum_{\underline{x}}a^{d-1}\left[\frac{1}{2}\left(\frac{\phi_{k+1}(\underline{x})-\phi_k(\underline{x})}{a}\right)^2+\frac{1}{2}(\underline{\nabla}\phi_k)^2(\underline{x})+\frac{1}{2}m^2\phi_k^2(\underline{x})\right]\right\}\Bigg|_{\substack{\phi_0\,=\,\varphi_{\rm i}\,,\\ \phi_n\,=\,\varphi_{\rm f}}} \tag{2.32}$$

In the zeroth and nth layer, ϕ is fixed to the initial and final values $\varphi_{\rm i}$ and $\varphi_{\rm f}$ respectively, while it is integrated over from layer 1 to $n-1$, weighted by the exponentiated action.

We can make this formula more symmetric with respect to the boundaries $t_{\rm i}$ and $t_{\rm f}$, if we add potential terms to the nth layer, writing

$$\left(\prod_{k=1}^{n-1}\prod_{\underline{x}}\int {\rm d}\phi_k(\underline{x})\right)\left(\prod_{\underline{x}}\left(\frac{a^{d-2}}{2\pi\hbar}\right)^{n/2}\right)\times$$
$$\exp\left\{-\frac{1}{\hbar}\sum_{\underline{x}}a^d\left[\sum_{k=0}^{n-1}\frac{1}{2}\left(\frac{\phi_{k+1}(\underline{x})-\phi_k(\underline{x})}{a}\right)^2+\sum_{k=0}^{n}\left(\frac{1}{2}(\underline{\nabla}\phi_k)^2(\underline{x})+\frac{1}{2}m^2\phi_k^2(\underline{x})\right)\right]\right\}\Bigg|_{\substack{\phi_0\,=\,\varphi_{\rm i}\,,\\ \phi_n\,=\,\varphi_{\rm f}}}. \tag{2.33}$$

Clearly, such a change does not affect the continuum limit. We also rewrite the normalization factors: in (2.33), there are n factors of

$$C_a := \sqrt{\frac{a^{d-2}}{2\pi\hbar}} \tag{2.34}$$

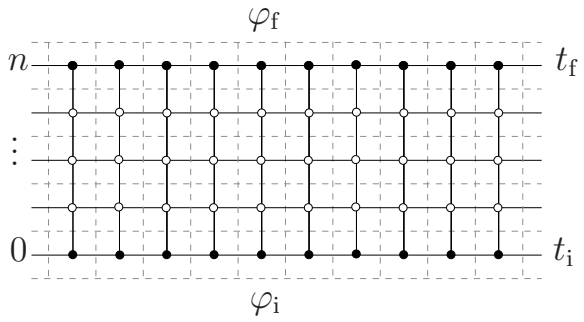


Figure 2.7: Lattice diagram for path integral on V_{fi}.

for each $\underline{x} \in S_a$. We express this in a more geometric fashion by attributing a factor C_a to every spacetime point $x = (\underline{x}, t_\mathrm{i} + ka)$ for which $\phi_k(\underline{x})$ is integrated over, and by associating a factor $\sqrt{C_a}$ to each point in the initial and final layer:

$$W_a[\varphi_\mathrm{f}, t_\mathrm{f}; \varphi_\mathrm{i}, t_\mathrm{i}] := \left(\prod_{\underline{x} \in S_a} \sqrt{C_a} \right)^2 \left(\prod_{k=1}^{n-1} \prod_{\underline{x} \in S_a} \int C_a \, \mathrm{d}\phi_k(\underline{x}) \right) \times$$
$$\exp\left\{ -\frac{1}{\hbar} \sum_{\underline{x} \in S_a} a^d \left[\sum_{k=0}^{n-1} \frac{1}{2} \left(\frac{\phi_{k+1}(\underline{x}) - \phi_k(\underline{x})}{a} \right)^2 + \sum_{k=0}^{n} \left(\frac{1}{2} (\underline{\nabla} \phi_k)^2(\underline{x}) + \frac{1}{2} m^2 \phi_k^2(\underline{x}) \right) \right] \right\} \Bigg|_{\substack{\phi_0 = \varphi_\mathrm{i}\,, \\ \phi_n = \varphi_\mathrm{f}}} . \tag{2.35}$$

In Fig. 2.7, this is represented diagrammatically for the case $d = 2$: open points stand for an integration over the associated field variable and carry a factor C_a. Boundary points are solid and contribute a factor $\sqrt{C_a}$. For each point there is a mass term in the action, and each link between points gives a term with the corresponding lattice derivative. The dual lattice is drawn shaded.

2.4.2 General Definition

By applying the same rules to more complicated arrangements of points, we can define a path integral regularization for general volumes V. Let $V \subset \mathbb{R}^d$ be open and its boundary Σ piecewise smooth. We use hypercubic lattices

$$L_a := \{ x \in a\mathbb{Z}^d \mid -aM \leq |x_\mu| \leq aM \,,\ \mu = 1, \ldots, d \}$$

with lattice constant $a > 0$ and edge length $2aM$, $M \in \mathbb{N}$. e_μ is the unit vector in the μth direction. A lattice point x and a direction μ define a link

$$l \equiv (x, \mu) \,.$$

The associated lattice gradient is

$$\nabla_l f \equiv \nabla_\mu f(x) := \frac{f(x + e_\mu) - f(x)}{a} \,.$$

Given a subset $P \subset L_a$, $l(P)$ denotes the set of links that connect points within P. Let

$$\tilde{V}_a := L_a \cap V$$

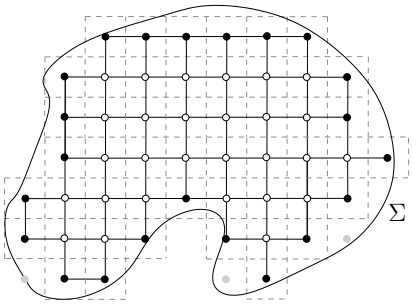


Figure 2.8: Lattice diagram for a general volume V.

be the intersection of V with the lattice. The points of $\tilde{V}_a$ fall into three categories (Fig. 2.8): we call a point *interior* if it has $2d$ links to points of $\tilde{V}_a$. If a point is linked to less than $2d$ points of $\tilde{V}_a$, but connected to at least one interior point, it is a *boundary point*. The remaining points of $\tilde{V}_a$ have only links to boundary points and we will not use them when representing the path integral on the lattice (they are drawn shaded in Fig. 2.8). The set of relevant points is therefore

$$V_a := I_a \cup \Sigma_a \,,$$

where I_a and Σ_a denote the set of interior and boundary points respectively.

On the lattice, the path integral becomes a summation over scalar fields $\phi : V_a \to \mathbb{R}$ on V_a. The action consists of contributions from links in $l(V_a)$ and points in V_a:

$$S[\phi, V_a] := \sum_{l \in l(V_a)} a^d \, \frac{1}{2} (\nabla_l \phi)^2 + \sum_{x \in V_a} a^d \, \frac{1}{2} m^2 \phi(x)$$

Given a continuous boundary field φ on Σ, one has to translate it into boundary data for V_a. We do so by defining the discrete boundary field

$$\varphi_a : \Sigma_a \to \mathbb{R} \,, \quad \varphi_a(x) = \varphi(\mathrm{pmd}_\Sigma(x)) \,,$$

The function pmd_Σ (pmd stands for "point of minimal distance") returns a point on Σ which has minimal distance to x. Now, we have all the necessary notation to give the regularized form of the propagator $W[\varphi, V]$: we specify it as

$$W_a[\varphi_a, V_a] := \left(\prod_{x \in \Sigma_a} \sqrt{C_a} \right) \left(\int \prod_{x \in I_a} C_a \, \mathrm{d}\phi(x) \right) \exp\left(-\frac{1}{\hbar} S[\phi, V_a] \right) \Big|_{\phi|_{\Sigma_a} = \varphi_a} \,, \tag{2.36}$$

with factors C_a as in (2.34). The continuum propagator $W[\varphi, V]$ is then defined by the limit of vanishing lattice constant and infinite lattice size:

$$W[\varphi, V] := \lim_{a \to 0} \lim_{M \to \infty} W_a[\varphi_a, V_a] \,.$$

To simplify notation, we omit the $\lim_{M \to \infty}$–symbol in the remainder of the text. That is, the limit of infinite lattice size (for constant a) is implicit in all subsequent formulas.

We now make a number of unproven assumptions about the regularization (2.36):

(A1) The propagator (2.36) has a continuum limit: that is, there is a non-trivial space $F(\Sigma)$ of boundary fields on Σ such that for each $\varphi \in F(\Sigma)$ the limit

$$W[\varphi, V] := \lim_{a \to 0} W_a[\varphi_a, V_a]$$

is well-defined.

(A2) W reproduces the conventional propagator: for $V_{\mathrm{fi}} = \mathbb{R}^{d-1} \times [t_{\mathrm{i}}, t_{\mathrm{f}}]$ and appropriate boundary conditions at spatial infinity,

$$W[(\varphi_{\mathrm{f}}, \varphi_{\mathrm{i}}), V_{\mathrm{fi}}] = \langle \varphi_{\mathrm{f}} | \mathrm{e}^{-H(t_{\mathrm{f}} - t_{\mathrm{i}})/\hbar} | \varphi_{\mathrm{i}} \rangle \,.$$

(A3) $W[\varphi, V]$ is translation and rotation invariant: i.e. under an isometry $f : \mathbb{R}^d \to \mathbb{R}^d$,

$$W[\varphi \circ f^{-1}, f(V)] = W[\varphi, V] \,.$$

(A4) There is a functional derivative $\dfrac{\delta}{\delta\varphi}$ on $F(\Sigma)$ whose action on $W[\varphi, V]$ can be approximated as follows:

$$\sum_{x \in \Sigma_a} a^{d-1} \frac{\partial^n W_a}{\partial (a^{d-1}\varphi_a(x))^n}[\varphi_a, V_a] \quad \stackrel{a \to 0}{\longrightarrow} \quad \int_\Sigma \mathrm{d}\Sigma(x) \frac{\delta^n W}{\delta \varphi(x)^n}[\varphi, V] \,.$$

To evaluate the path integral (2.36), it is useful to arrange the field variables from each point in vectors ϕ and write the action as

$$S[\phi, V_a] = \frac{1}{2} \phi \cdot B_a \, \phi + c_a \cdot \phi + d_a \,. \tag{2.37}$$

The boundary fields φ_a are contained in the vectors c_a and d_a respectively. The action is bounded from below, so for each φ_a, there is at least one solution ϕ_{cl} of the Euclidean equations of motion

$$\frac{\partial S}{\partial \phi}[\varphi_a, V_a] = B_a \, \phi + c_a = 0 \,.$$

If B_a is non-degenerate, the solution is unique and one can define the Hamilton function

$$S[\varphi_a, V_a] := S[\phi_{\mathrm{cl}}, V_a]$$

for the discrete Euclidean system. We assume, in fact, that

(A5) The matrix B_a is non-degenerate and the Hamilton function $S[\varphi_a, V_a]$ is analytic in φ_a.

The change of variables $\xi := \phi - \phi_{\mathrm{cl}}$ renders the action (2.37) quadratic:

$$S[\xi, V_a] = \frac{1}{2} \xi \cdot B_a \, \xi + S[\varphi_a, V_a]$$

The integral (2.36) becomes Gaussian and gives

$$W_a[\varphi_a, V_a] = \left(\prod_{x \in \Sigma_a} \sqrt{C_a} \right) \left(\int \prod_{x \in I_a} \sqrt{2\pi} \, C_a \right) \frac{1}{\sqrt{\det B_a}} \exp\left(-\frac{1}{\hbar} S[\varphi_a, V_a] \right) \,. \tag{2.38}$$

Therefore, by **(A5)**, the regularized kernel $W_a[\varphi_a, V_a]$ must be analytic in φ_a, which will be used in section 2.5.1 when deriving the Schrödinger equation.

Remark: In **(A1)** and **(A5)** we have formulated the continuum limit in terms of pointwise convergence, i.e. by separate convergence for each boundary field φ in $F(\Sigma)$. According to (2.38), the field dependence of the regularized kernel resides only in the Hamilton function $S[\varphi_a, V_a]$. The latter converges against the continuum function $S[\varphi, V]$, which is defined pointwise. Thus, it is plausible to assume that the continuum propagator $W[\varphi, V]$, too, is a pointwise function on $F(\Sigma)$. When further developing the formalism, pointwise convergence is likely to be replaced by convergence in a Hilbert space norm or other measures which only distinguish between equivalence classes of boundary fields. For the purpose of this article, however, it is sufficient and simplifies notation if we use convergence on single fields.

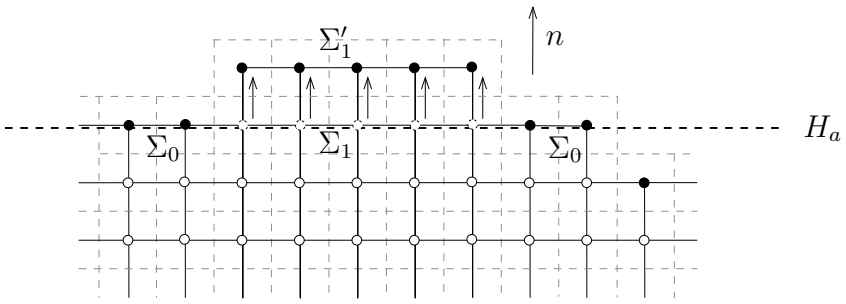


Figure 2.9: Addition of a single layer.

2.5 Generalized Schrödinger Equation

The propagator W depends on a spacetime region V and a field φ specified on the boundary Σ. Thus, as for the Hamilton function in section 2.3, one can define a deformation derivative: using the same notation as there, we set

$$L_N W[\varphi, V] := \lim_{s\to 0} \frac{W[\varphi^s, V^s] - W[\varphi, V]}{s} \,.$$

In this section, we derive that

$$\hbar L_N W[\varphi, V] = \left(-H_N[\varphi, \hbar\frac{\delta}{\delta\varphi}, V] + P_N[\varphi, \hbar\frac{\delta}{\delta\varphi}, V] \right) W[\varphi, V] \,, \tag{2.39}$$

where

$$\begin{aligned} H_N[\varphi, \pi, V] &:= \int_\Sigma \mathrm{d}\Sigma\; N_\perp \left(-\frac{1}{2}\pi^2 + \frac{1}{2}(\nabla_\Sigma \varphi)^2 + \frac{1}{2} m^2 \varphi^2 \right) \,, \\ P_N[\varphi, \pi, V] &:= -\int_\Sigma \mathrm{d}\Sigma\, N_\parallel \cdot \nabla_\Sigma \varphi\, \pi \,. \end{aligned}$$

When $V = \mathbb{R}^{d-1} \times [t_\mathrm{i}, t_\mathrm{f}]$ and $N_\perp|_{t_\mathrm{f}} = 1$, $N_\perp|_{t_\mathrm{i}} = 0$, $N_\parallel = 0$, this yields the ordinary Euclidean Schrödinger equation

$$\left(\hbar\frac{\partial}{\partial t_\mathrm{f}} + H[\varphi_\mathrm{f}, \hbar\frac{\delta}{\delta\varphi_\mathrm{f}}] \right) W[\varphi_\mathrm{f}, t_\mathrm{f}; \varphi_\mathrm{i}, t_\mathrm{i}] = 0 \,.$$

The strategy of the derivation: using assumption **(A5)** and rotation invariance **(A3)**, we show that the regularized propagator satisfies a lattice version of equation (2.39) when V is deformed along flat parts of its boundary. The central step is analogous to the calculation Feynman used when deriving the Schrödinger equation from the path integral of a point particle [49] (see also chap. 4, [52]). Due to **(A1)** and **(A4)**, the discrete equations have the continuum limit (2.39). To cover also the case, when deformations are applied to curved sections of Σ, we approximate Σ by a triangulation, apply (2.39) to each triangle and let the fineness of the triangulation go to zero.

For simplicity, the argument is formulated for *bounded* volumes below. The generalization to infinitely extended V is straightforward.

2.5.1 Discrete Schrödinger Equation

Consider a lattice diagram in which part of Σ_a coincides with a hypersurface H_a of the lattice L_a. Let n denote the normal vector of H_a. The simplest way of modifying such a diagram is

to add a $(d-1)$-dimensional layer of points along $H_a \cap \Sigma_a$ (see Fig. 2.9). The old boundary points adjacent to the layer become interior points. We describe this by a lapse function $N_a : \Sigma_a \to \{0,1\}$ which indicates for any given point of the boundary if a new point will be linked to it or not. Then, the function

$$\sigma_a : \Sigma_a \to L_a\,, \quad x \mapsto x + aN_a(x)\,n$$

is the discrete flow associated to the deformation of the boundary Σ_a. Define the new diagram and its boundary by

$$V_a' := V_a \cup \sigma_a(\Sigma_a)\,, \qquad \Sigma_a' := \sigma_a(\Sigma_a)\,.$$

The set

$$\Sigma_1 := N_a^{-1}(1)$$

is the part of Σ_a which is moved and becomes $\Sigma_1' := \sigma_a(\Sigma_1)$, while

$$\Sigma_0 := N_a^{-1}(0)$$

remains unchanged. As in the continuous case, we choose the new boundary field to be the pull-forward of the old one, that is,

$$\varphi_a' := \varphi_a \circ \sigma_a^{-1}\,.$$

The resulting path integral is

$$\begin{aligned} W_a[\varphi_a', V_a'] = {} & \left(\prod_{x\in\Sigma_0\cup\Sigma_1'} \sqrt{C_a}\right)\left(\prod_{x\in I_a\cup\Sigma_1}\int C_a\,\mathrm{d}\phi(x)\right)\times \\ & \times \exp\left\{-\frac{a^d}{\hbar}\left[\sum_{x\in\Sigma_1}\frac{1}{2}\left(\frac{\varphi_a'(\sigma_a(x))-\phi(x)}{a}\right)^2 + \sum_{l\in l(\Sigma_1')}\frac{1}{2}(\nabla_l\varphi_a')^2 + \sum_{x\in\Sigma_1'}\frac{1}{2}m^2{\varphi_a'}^2(x)\right]\right. \\ & \left. - \frac{1}{\hbar}S[\phi, V_a]\right\}\Big|_{\phi|_{\Sigma_0}=\varphi_a|_{\Sigma_0}} \end{aligned}$$

(Recall that $l(\Sigma_1')$ is the set of links between points of Σ_1'.) The same can also be written as a convolution with the original propagator:

$$\begin{aligned} & \left(\prod_{x\in\Sigma_1}\int C_a\,\mathrm{d}\phi(x)\right)\exp\left\{-\frac{a^d}{\hbar}\left[\sum_{x\in\Sigma_1}\frac{1}{2}\left(\frac{\varphi_a(x)-\phi(x)}{a}\right)^2\right.\right. \\ & \qquad \left.\left. + \sum_{l\in l(\Sigma_1)}\frac{1}{2}(\nabla_l\varphi_a)^2 + \sum_{x\in\Sigma_1}\frac{1}{2}m^2\varphi_a^2(x)\right]\right\} W_a[(\varphi_a|_{\Sigma_0},\phi), V_a] \end{aligned}$$

Following Feynman's derivation of the Schrödinger equation, we introduce new variables

$$\xi(x) := \sqrt{\frac{a^{d-2}}{\hbar}}(\phi(x) - \varphi_a(x))\,, \quad x \in \Sigma_1$$

and get

$$W_a[\varphi_a', V_a'] \;=\; \left(\prod_{x\in\Sigma_1}\int \overbrace{C_a\sqrt{\frac{\hbar}{a^{d-2}}}}^{=1/\sqrt{2\pi}}\,\mathrm{d}\xi(x)\right)$$

$$\times \exp\left\{-\frac{1}{2}\sum_{x\in\Sigma_1}\frac{\xi^2(x)}{a} - \frac{a^d}{\hbar}\left[\sum_{l\in l(\Sigma_1)}\frac{1}{2}(\nabla_l\varphi_a)^2 + \sum_{x\in\Sigma_1}\frac{1}{2}m^2\varphi_a^2(x)\right]\right\}$$
$$\times\ W_a\left[\left(\varphi_a|_{\Sigma_0}, \varphi_a|_{\Sigma_1} + \sqrt{\hbar/a^{d-2}}\,\xi\right),\, V_a\right]$$

Next we apply Laplace's method to obtain an asymptotic expansion of this expression (see e.g. chap. 11, [52]): the dominant contribution to the Gaussian integral comes from an a-dependent interval $[-\varepsilon_a, \varepsilon_a]^{|\Sigma_1|}$ around $\xi = 0$. ($|\Sigma_1|$ denotes the number of points in Σ_1.) The integral outside is exponentially damped for $a \to 0$ and neglected. Within the interval, one can Taylor expand W_a in ξ and reverse the order of integration and Taylor expansion. To evaluate the integration for each term, the integration range is extended back to its full size: this introduces an error in each term of the sum and convergence is lost, but the expansion is still valid asymptotically for $a \to 0$.

For the integrations and estimates, we use the following formulas:

$$\int_{-\infty}^{\infty} \mathrm{d}y\, y^n\, \mathrm{e}^{-y^2/2} = \begin{cases} (n-1)(n-3)\cdots 3\cdot 1\cdot\sqrt{2\pi} & ,\ n \geq 0 \text{ and even}, \\ 0 & ,\ n \text{ odd}, \end{cases} \tag{2.40}$$

$$\int_{\pm\varepsilon_a}^{\pm\infty} \mathrm{d}y\, y^n\, \mathrm{e}^{-y^2/2} = O(\varepsilon_a^{n-1}\,\mathrm{e}^{-\varepsilon_a^2/2}) \quad \text{as } a \to 0\,. \tag{2.41}$$

We set $\epsilon_a = 1/a$. Consider first the integral outside the chosen interval:

$$\left(\prod_{x\in\Sigma_1}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}\setminus[-1/a,1/a]}\mathrm{d}\xi(x)\right)\exp\left(-\frac{1}{2}\sum_{x\in\Sigma_1}\frac{\xi^2(x)}{a}\right)$$
$$\times\ \exp\left\{-\frac{a^d}{\hbar}\left[\sum_{l\in l(\Sigma_1)}\frac{1}{2}(\nabla_l\varphi_a)^2 + \sum_{x\in\Sigma_1}\frac{1}{2}m^2\varphi_a^2(x)\right]\right\}$$
$$\times\ W_a\left[\left(\varphi_a|_{\Sigma_0}, \varphi_a|_{\Sigma_1} + \sqrt{\hbar/a^{d-2}}\,\xi\right),\, V_a\right] \tag{2.42}$$

The second exponent vanishes in the continuum limit. For W_a, we employ formula (2.38) and replace $\exp(-S[\ldots, V_a])$ by 1, as the action is positive. The determinant and C_a-factors are together of order $O(1)$, since, by assumption, (2.38) approaches a finite continuum limit. Thus, the modulus of (2.42) is smaller than

$$\left(\prod_{x\in\Sigma_1}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}\setminus[-1/a,1/a]}\mathrm{d}\xi(x)\,\mathrm{e}^{-\xi^2(x)/2}\right)\cdot O(1) \overset{(2.41)}{\leq} O\left(\frac{1}{a}\,\mathrm{e}^{-\frac{|\Sigma_1|}{2a^2}}\right) \quad \text{as } a \to 0\,.$$

In the integral over $[-1/a, 1/a]^{|\Sigma_1|}$, we Taylor expand W_a in ξ:

$$\left(\prod_{x\in\Sigma_1}\frac{1}{\sqrt{2\pi}}\int_{-1/a}^{1/a}\mathrm{d}\xi(x)\right)\exp\left(-\frac{1}{2}\sum_{x\in\Sigma_1}\frac{\xi^2(x)}{a}\right)$$

$$\times \exp\left\{-\frac{a^d}{\hbar}\left[\sum_{l\in l(\Sigma_1)}\frac{1}{2}(\nabla_l\varphi_a)^2+\sum_{x\in\Sigma_1}\frac{1}{2}m^2\varphi_a^2(x)\right]\right\}$$
$$\times\left(W_a[\varphi_a,V_a]+\sum_{x\in\Sigma_1}\frac{\partial W_a}{\partial\varphi_a(x)}\sqrt{\frac{\hbar}{a^{d-2}}}\,\xi(x)+\frac{1}{2}\sum_{x,y\in\Sigma_1}\frac{\partial^2 W_a}{\partial\varphi_a(x)\partial\varphi_a(y)}\,\frac{\hbar}{a^{d-2}}\,\xi(x)\xi(y)+\dots\right)$$

By assumption **(A5)**, W_a is analytic in the field variable, so the Taylor expansion converges uniformly and we are allowed to integrate each term of the series separately. We also set the limits of integration back to plus and minus infinity. This does not affect the *asymptotic* property of the series, since for each term the resulting error is only exponentially small: for example, the linear term yields

$$\left|\left(\prod_{x\in\Sigma_1}\frac{1}{\sqrt{2\pi}}\int_{-1/a}^{1/a}\mathrm{d}\xi(x)\right)\mathrm{e}^{-\xi^2(x)/2}\sum_{x\in\Sigma_1}a^{d-1}\frac{\partial W_a}{\partial(a^{d-1}\varphi_a(x))}\sqrt{\frac{\hbar}{a^{d-2}}}\,\xi(x)\right|\le O\left(\frac{|\Sigma_1|}{\sqrt{a^{d-2}}}\,\mathrm{e}^{-\frac{1}{2a^2}}\right),$$

because of (2.41) and **(A4)**. Then, we can use equation (2.40) to do the Gaussian integration in each term of the asymptotic series. Each integration, that is, each point $x\in\Sigma_1$, leaves an overall factor $\sqrt{2\pi}$. Terms with an uneven number of ξ variables (of the same point) vanish. We obtain

$$W_a[\varphi'_a,V'_a]\;\sim\;\exp\left\{-\frac{a^d}{\hbar}\left[\sum_{l\in l(\Sigma_1)}\frac{1}{2}(\nabla_l\varphi_a)^2+\sum_{x\in\Sigma_1}\frac{1}{2}m^2\varphi_a^2(x)\right]\right\}$$
$$\times\left(W_a[\varphi_a,V_a]+\sum_{x\in\Sigma_1}\frac{\hbar}{2}\frac{\partial^2 W_a}{\partial\varphi_a(x)^2}\cdot\frac{1}{a^{d-2}}+\sum_{n=2}^{\infty}\sum_{x\in\Sigma_1}c(n)\frac{\partial^{2n}W_a}{\partial\varphi_a(x)^{2n}}\cdot\frac{1}{(a^{d-2})^n}\right),$$

where the $c(n)$'s are numerical coefficients. If we write φ_a as the pull-back $\varphi'_a\circ\sigma_a\equiv\sigma_a^*\varphi'_a$ of φ'_a and use also the lapse function N_a, the final result becomes

$$W_a[\varphi'_a,V'_a]\;\sim\;\exp\left\{-\frac{1}{\hbar}\left[\sum_{l\in l(\Sigma_a)}a^{d-1}\,aN_a\,\frac{1}{2}(\nabla_l\sigma_a^*\varphi'_a)^2+\sum_{x\in\Sigma_a}a^{d-1}\,aN_a\frac{1}{2}m^2(\sigma_a^*\varphi'_a)^2(x)\right]\right\}$$
$$\times\left\{W_a[\sigma_a^*\varphi'_a,V_a]+\sum_{x\in\Sigma_a}a^{d-1}\,aN_a\left(\frac{\hbar}{2}\frac{\partial^2 W_a}{\partial(a^{d-1}\varphi_a(x))^2}[\sigma_a^*\varphi'_a,V_a]\right.\right.$$
$$\left.\left.+\sum_{n=2}^{\infty}a^{d(n-1)}\,c(n)\frac{\partial^{2n}W_a}{\partial(a^{d-1}\varphi_a(x))^{2n}}[\sigma_a^*\varphi'_a,V_a]\right)\right\}. \tag{2.43}$$

Iteration

Suppose now that the deformed set V'_a does not arise from the addition of a single layer, but from a continuous deformation V^s of the orginial volume V. That is, we want to compare $W_a[\varphi_a^s,V_a^s]$ against $W_a[\varphi_a,V_a]$. To make the calculation tractable, we require that the vector field N vanishes outside a neighbourhood U of a boundary point $x\in\Sigma$, and that within U the boundary Σ is flat. Denote this part of Σ by $\Sigma_U:=\Sigma\cap U$.

By rotation and translation invariance (**(A3)**), we can orient V such that Σ_U coincides with a hyperplane of the lattice L_a. Let us begin by considering the case where the lapse $N_\perp$ is positive, that is, $V\subset V^s$. For small enough s, the typical diagram for $W_a[\varphi_a^s,V_a^s]$ looks like Fig. 2.10 (or its higher-dimensional equivalent), where along the normal direction n each point of Σ_a is in one-to-one correspondence with a point of Σ_a^s. (Note that in the limit $s\to 0$, the slope of Σ^s against Σ becomes arbitrarily small.) The new boundary Σ_a^s

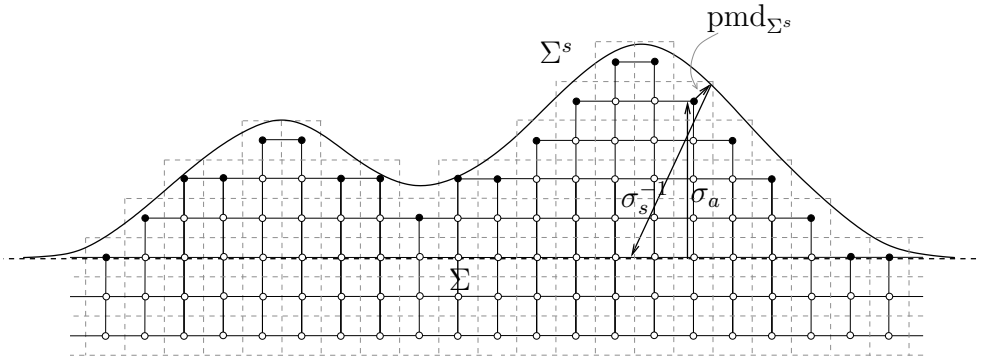


Figure 2.10: Diagram for V_a^s.

can be built from Σ_a by repeatedly adding single layers as described previously. Thus, we can iterate formula (2.43) to obtain a relation between $W_a[\varphi_a^s, V_a^s]$ and $W_a[\sigma_a^*\varphi_a^s, V_a]$ where now, σ_a is the concatenation of all single-step flows. When collecting the various terms of the iteration, the lapse functions for each step add up to the *total* lapse function N_a. We order the result in powers of aN_a and a:

$$\begin{aligned} W_a[\varphi_a^s, V_a^s] &= W_a[\sigma_a^*\varphi_a^s, V_a] + \sum_{x\in\Sigma_a} a^{d-1}\, aN_a \frac{\hbar}{2} \frac{\partial^2 W_a}{\partial(a^{d-1}\varphi_a(x))^2}[\sigma_a^*\varphi_a^s, V_a] \\ &\quad - \frac{1}{\hbar}\left[\sum_{l\in l(\Sigma_a)} a^{d-1}\, aN_a\, \frac{1}{2}(\nabla_l \sigma_a^*\varphi_a^s)^2 + \sum_{x\in\Sigma_a} a^{d-1}\, aN_a \frac{1}{2} m^2 (\sigma_a^*\varphi_a^s)^2(x)\right] \\ &\quad + O(a^d(aN_a)) + O((aN_a)^2) \qquad \text{as } a, s \to 0\,. \end{aligned} \tag{2.44}$$

Note that the displacement vector aN_a approaches $sN_\perp$ when both a and s become small, i.e.

$$aN_a = sN_\perp + O(s^2) + O(a)\,. \tag{2.45}$$

If N is normal to Σ_U, the discrete flow σ_a approximates the continuous one, σ_s, and

$$\begin{aligned} \sigma_a^*\varphi_a^s &= \varphi_a^s \circ \sigma_a = \varphi^s \circ \mathrm{pmd}_{\Sigma_s} \circ \sigma_a \\ &= \varphi \circ \sigma_s^{-1} \circ \mathrm{pmd}_{\Sigma_s} \circ \sigma_a \\ &= \varphi_a + O(a)\,. \end{aligned}$$

In general, N has also a tangential component, so

$$\sigma_a^*\varphi_a^s = \varphi_a - sN_\parallel^\mu \nabla_\mu \varphi_a + O(s^2) + O(a)\,, \tag{2.46}$$

as can be seen from the arrow diagram in Fig. 2.10. Plugging (2.45) and (2.46) into (2.44), one arrives at a regularized form of the Euclidean Schrödinger equation:

$$\frac{W_a[\varphi_a^s, V_a^s] - W_a[\varphi_a, V_a]}{s} = \hat{O}_a W_a[\varphi_a, V_a] + O(s) + O(a) + O(a/s)\,, \tag{2.47}$$

where $\hat{O}_a$ is the operator

$$\begin{aligned} \hat{O}_a &:= -\frac{1}{\hbar}\sum_{x\in\Sigma_a} a^{d-1}\left[N_\perp(x)\left(-\frac{\hbar^2}{2}\frac{\partial^2}{\partial(a^{d-1}\varphi_a(x))^2} + \frac{1}{2}m^2\varphi_a^2(x)\right)\right. \\ &\quad \left. + N_\parallel^\mu(x)\nabla_\mu\varphi_a(x)\,\hbar\frac{\partial}{\partial(a^{d-1}\varphi_a(x))}\right] - \frac{1}{\hbar}\sum_{l\in l(\Sigma_a)} a^{d-1} N_\perp(x)\,\frac{1}{2}(\nabla_l\varphi_a)^2 \end{aligned} \tag{2.48}$$

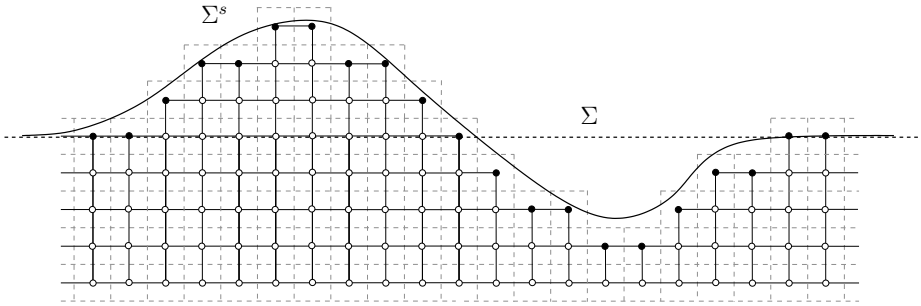


Figure 2.11: Lapse with positive and negative sign.

An analogous argument applies to the case of negative lapse $N_\perp$. For mixed diagrams as in Fig. 2.11, both types of calculations can be combined to give (2.47) for lapses of arbitrary sign.

2.5.2 Continuous Schrödinger Equation

Choose N as before, i.e. with support on a neighbourhood U of $x \in \Sigma$ where $\Sigma \cap U$ is flat. We want to show that

$$L_N W[\varphi, V] = \lim_{s\to 0} \frac{W[\varphi^s, V^s] - W[\varphi, V]}{s} = \hat{O} W[\varphi, V]\,, \tag{2.49}$$

where

$$\begin{aligned} \hat{O} \;:=\; -\frac{1}{\hbar}\int_\Sigma \mathrm{d}\Sigma(x) \Big[N_\perp(x) \Big(-\frac{\hbar^2}{2}\frac{\delta^2}{\delta\varphi(x)^2} + \frac{1}{2}(\nabla_\Sigma\varphi)^2(x) + \frac{1}{2}m^2\varphi^2(x) \Big) \\ + N_\parallel(x)\cdot\nabla_\Sigma\varphi(x)\,\hbar\frac{\delta}{\delta\varphi(x)} \Big]\,. \end{aligned}$$

Stated more explicitly, this means that for any $\epsilon > 0$ there is an $s_0 > 0$ such that

$$\left| \frac{W[\varphi^s, V^s] - W[\varphi, V]}{s} - \hat{O} W[\varphi, V] \right| \;\le\; \epsilon \qquad \text{for all } s < s_0\,. \tag{2.50}$$

To obtain an upper estimate on the left-hand side, we insert regularized propagators and operators in a suitable way, and then apply the triangle inequality:

LHS of (2.50)

$$\begin{aligned} &= \Big| \frac{1}{s}\Big(W[\varphi^s, V^s] - W_a[\varphi_a^s, V_a^s] + W_a[\varphi_a^s, V_a^s] - W_a[\varphi_a, V_a] + W_a[\varphi_a, V_a] - W[\varphi, V] \Big) \\ &\quad - \hat{O}_a W_a[\varphi_a, V_a] + \hat{O}_a W_a[\varphi_a, V_a] - \hat{O} W[\varphi, V] \Big| \\ &\le \frac{1}{s}\Big| W[\varphi^s, V^s] - W_a[\varphi_a^s, V_a^s] \Big| + \frac{1}{s}\Big| W[\varphi, V] - W_a[\varphi_a, V_a] \Big| \\ &\quad + \Big| \hat{O} W[\varphi, V] - \hat{O}_a W_a[\varphi_a, V_a] \Big| \\ &\quad + \left| \frac{W_a[\varphi_a^s, V_a^s] - W_a[\varphi_a, V_a]}{s} - \hat{O}_a W_a[\varphi_a, V_a] \right|. \end{aligned}$$

By assumption **(A1)** (existence of the continuum limit), the first two terms become smaller than $\epsilon/4$ when the lattice constant a is smaller than some $a_s > 0$. The partial derivatives

and potential terms in (2.48) approach their continuum analogues as $a \to 0$, so there is also an $a_0 > 0$ such that

$$\left|\hat{O}W[\varphi, V] - \hat{O}_a W_a[\varphi_a, V_a]\right| < \frac{\epsilon}{4} \qquad \text{for all } a < a_0 \,.$$

The regularized Schrödinger equation tells us that for s smaller than some s_0, there is an $a'_s > 0$ such that

$$\left|\frac{W_a[\varphi_a^s, V_a^s] - W_a[\varphi_a, V_a]}{s} - \hat{O}_a W_a[\varphi_a, V_a]\right| < \frac{\epsilon}{4} \qquad \text{for all } a < a'_s \,.$$

Thus, for any $s < s_0$, we can choose $a < \min\{a_s, a_0, a'_s\}$ and the left-hand side of (2.50) must be smaller than ϵ. □

2.5.3 Curved Boundaries

As it is based on the lattice equation (2.47), the previous derivation applies only when flat sections of the boundary Σ are deformed. We do not know how to extend the lattice calculation to the case where both initial and deformed surface are curved. Below we give an argument which circumvents this difficulty, but requires additional assumptions. The idea is to approximate the curved boundary by a triangulation, apply the variation to each of the flat triangles and add up the contributions.

Let T_δ be a triangulation of Σ with fineness δ: that is, when two 0-simplices are connected by a 1-simplex, their metric distance is at most δ. Let $\{\Sigma_\alpha\}$ denote the set of $(d-1)$-simplices $\Sigma_\alpha \subset \Sigma$ of the triangulation. The corner points of each such simplex Σ_α define a $(d-1)$-simplex in $\mathbb{R}^d$ which we call Σ_{Δ_α}. The hypersurface $\Sigma_\Delta := \cup_\alpha \Sigma_{\Delta_\alpha}$ approximates Σ and encloses the volume V_Δ. We can view V as a deformation of V_Δ and find a flow

$$\rho : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R} \,, \quad (t, x) \mapsto \rho(t, x) \equiv \rho_t(x)$$

such that $\rho_1(V_\Delta) = V$ and $\rho_1(\Sigma_{\Delta_\alpha}) = \Sigma_\alpha$. We equip Σ_Δ with the boundary field $\varphi_\Delta := \rho_1^* \varphi = \varphi \circ \rho_1$, the pull-back of φ under this flow. Motivated by equation (2.49) for flat surfaces, we assume that the difference between $W[\varphi, V]$ and $W[\varphi_\Delta, V_\Delta]$ is of the order of the volume difference between V and V_Δ:

$$\begin{aligned} W[\varphi, V] &= W[\rho_{1*}\varphi_\Delta, \rho_1(V_\Delta)] \\ &= W[\varphi_\Delta, V_\Delta] + O(|V - V_\Delta|) \,. \end{aligned} \tag{2.51}$$

Next we introduce "characteristic" functions $\chi_\alpha : \mathbb{R}^d \to \mathbb{R}$ with the property that

$$\begin{aligned} &\chi_\alpha(x) = 1 \quad \text{for} \quad x \in \Sigma_{\Delta_\alpha} \,, \\ &\chi_\alpha(x) = 0 \quad \text{for} \quad x \in \Sigma_{\Delta_\beta} \,,\ \alpha \neq \beta \,, \\ \text{and} \quad &\textstyle\sum_\alpha \chi_\alpha(x) = 1 \ \text{ for all } \ x \in \mathbb{R}^d \,. \end{aligned}$$

Using these functions, we can decompose the deformation field N according to

$$N = \sum_\alpha \chi_\alpha N \equiv \sum_\alpha N_\alpha \,.$$

Each component N_α is a discontinuous vector field and gives rise to a discontinuuous flow within $\mathbb{R}^d$. Suppose that by a suitable limiting procedure, one can define L_{N_α} such that equation (2.49) holds and

$$L_N = \sum_\alpha L_{N_\alpha} \,.$$

Then, equation (2.51) becomes

$$L_N W[\varphi, V] = \sum_\alpha L_{N_\alpha} W[\varphi_\Delta, V_\Delta] + O(|V - V_\Delta|)\,.$$

By construction, the vector fields N_α are only nonzero on the flat simplices Σ_{Δ_α}. Therefore, our result for flat surfaces (equation (2.49)) is applicable and yields

$$\begin{aligned}
&L_N W[\varphi, V] \\
&= \sum_\alpha \left\{ -\frac{1}{\hbar} \int_\Sigma \mathrm{d}\Sigma \left[N_{\alpha\perp} \left(-\frac{\hbar^2}{2} \frac{\delta^2}{\delta\varphi^2} + \frac{1}{2} (\nabla_\Sigma \varphi_\Delta)^2 + \frac{1}{2} m^2 \varphi_\Delta^2 \right) + N_\parallel \cdot \nabla_\Sigma \varphi_\Delta \, \hbar \frac{\delta}{\delta\varphi} \right] W[\varphi_\Delta, V_\Delta] \right\} \\
&\quad + O(|V - V_\Delta|) \\
&= -\frac{1}{\hbar} \int_\Sigma \mathrm{d}\Sigma \left[N_\perp \left(-\frac{\hbar^2}{2} \frac{\delta^2}{\delta\varphi^2} + \frac{1}{2} (\nabla_\Sigma \varphi)^2 + \frac{1}{2} m^2 \varphi^2 \right) + N_\parallel \cdot \nabla_\Sigma \varphi \, \hbar \frac{\delta}{\delta\varphi} \right] W[\varphi, V] \\
&\quad + O(|V - V_\Delta|)
\end{aligned}$$

In the $\delta \to 0$ limit, $|V - V_\Delta|$ goes to zero and one recovers the generalized Schrödinger equation for curved boundaries.

2.6 Summary and Discussion

We have proposed an exact definition for a Euclidean free scalar propagator $W[\varphi, V]$ which "evolves" wavefunctionals of fields along general spacetime domains V. Our main result is a derivation of the evolution equation

$$\hbar L_N W[\varphi, V] = \left(-H_N[\varphi, \hbar \frac{\delta}{\delta\varphi}, V] + P_N[\varphi, \hbar \frac{\delta}{\delta\varphi}, V] \right) W[\varphi, V]\,. \tag{2.52}$$

This equation describes how $W[\varphi, V]$ varies under infinitesimal deformations of V generated by a vector field N. The variation is given by the action of two operators: one is related to the field Hamiltonian and arises from normal deformations of the boundary $\Sigma = \partial V$. The second operator results from tangential deformations and generalizes the field momentum.

We showed also that the Hamilton function of the classical system satisfies an analogous Hamilton-Jacobi equation

$$L_N S[\varphi, V] = H_N[\varphi, \frac{\delta S}{\delta\varphi}, V] + P_N[\varphi, \frac{\delta S}{\delta\varphi}, V]\,. \tag{2.53}$$

When the boundary Σ consists of two infinite hyperplanes at fixed times, (2.52) and (2.53) reduce to the standard Schrödinger and Hamilton-Jacobi equation in their Euclidean form.

The derivation of eq. (2.52) is based on assumptions which we consider plausible, but are not proven. Most importantly, we have not shown that the proposed regularization of the propagator has a well-defined continuum limit. A description for converting the Euclidean to a Lorentzian propagator is missing. As described in section 2.2, we expect that an evolution equation analogous to (2.52) holds also for Lorentzian propagators. We emphasize that such state evolution may, in general, be non-unitary and nevertheless admit a physical interpretation.

Let us mention that eq. (2.52) is confirmed by a formal derivation by Doplicher that is based on the use of Green's functions [53]: the kernel W is formally expressed as

$$W[\varphi, V] = \sqrt{\det G_V(x, y)}\; S[\varphi, V] \tag{2.54}$$

where G_V satisfies

$$\left(\Box - m^2\right) G_V(x,y) = \delta(x-y)\,, \quad x,y \in V\,, \qquad G_V(x,y) = 0\,, \quad x \in \Sigma\,. \tag{2.55}$$

The derivative $L_N W$ is computed with the help of the Hadamard formula which gives the variation of G_V under deformations of V.

Chapter 3

Free vacuum for loop quantum gravity

3.1 Introduction

In the research on canonical loop quantum gravity (LQG) and spin foam models, one of the major open problems is the development of a reliable semiclassical analysis. At present, we do not know if these theories contain semiclassical states which correctly reproduce the observed Einstein gravity. Nor is a perturbative expansion available that would allow one to calculate the scattering of low-energy excitations of such states. Several lines of research have led to proposals for vacuum states, tentative results on perturbations, and modified dispersion relations for matter: among them are approaches based on the Kodama state [54]-[55], spin network invariants [56, 57], linearized gravity [58, 59], coherent states [60]-[61], weaves [62]-[63] and general boundaries [10]. While the kinematics of the gravitational sector and the dynamics of the matter sector are relatively well understood, the gravitational dynamics and its semiclassical limit remain largely unclear. Consequently, we have no conclusive evidence that loop quantum gravity has a physically realistic semiclassical and low-energy limit[1].

The work of this article is motivated by the following question: can one obtain a semiclassical perturbation series for LQG and, if yes, how? At first sight, it may seem odd to ask this question about a theory that has often been characterized as the *non-perturbative* alternative to perturbative approaches in quantum gravity. It is the very failure of traditional perturbation theory that led to the loop approach to quantum gravity, and one of its strongest points is the fact that it does not rely on any approximative scheme for its definition. On the other hand, we know that theories can have both perturbative and non-perturbative regimes, depending on which scale we look at them. In QED, for example, perturbative expansions work fine at low-energies where the fine-structure constant is small, while at higher energies the coupling running grows and perturbation theory breaks down. It is renormalization that relates the different regimes of the same theory, and allows us, in principle, to compute low-energy actions from actions at more fundamental scales. It gives us, in particular, the relation of measurable couplings at accessible scales to bare couplings at cutoff scales.

When the transition from one scale to another involves strong coupling regimes, the renormalization procedure may require non-perturbative techniques. Thus, it can happen that a field theory is perturbative at low energies, while a *non-perturbative* renormalization is needed to compute its low-energy action. This suggests that the failure of perturbative gravity may not come from the coupling expansion itself, but from the *perturbative* renormalization that fails to provide us with unique couplings in the effective action. Such a viewpoint is supported by the work of Reuter and others (see e.g. [64]-[65]) who investigate non-perturbative renormalization group flows of gravitational actions.

[1] For an overview on the literature, see sec. II.3, [12].

>From that perspective, it is conceivable that loop quantum gravity is a non-perturbative theory at its fundamental scale, and nevertheless accessible to perturbative treatments at lower scales. While possible in principle, this scenario remains rather elusive, since so far the loop formalism does not dispose of the techniques to implement renormalization and perturbation theory. Tentative ideas on renormalization have been formulated by Markopoulou and Oeckl [66, 67], and there are first attempts to develop a perturbation theory around the Kodama state [68, 55]. Starodubtsev constructs string-like excitations of this state [69], and Smolin has analyzed string perturbations [70] of causal spin networks [71, 72].

Here, we will take a more conservative strategy and try to proceed in close analogy with ordinary quantum field theory. There, the working principle of perturbation theory can be roughly described as follows: we choose a classical background and consider only states which are semiclassically peaked around it. As a result, the Hamiltonian is dominated by the lowest order terms in the fluctuation—the free part—whereas higher orders can be treated as small corrections. The free Hamiltonian defines a linear system and provides a first approximation to the dynamics of the perturbative regime. In spite of its simplicity, it is far from clear how this scheme should be transferred to the framework of loop quantum gravity: how can we generalize it to a theory, whose quantum degrees of freedom are distributional fields with 1-dimensional support—or dually to it—labelled networks? What is the analogue of a field fluctuation in a space of networks? What tells us how to separate an operator on graphs into a "free" and "interaction" part?

In the present paper, we approach this problem by constructing candidate states for a free vacuum and free gravitons in loop quantum gravity. We analyze these states in the hope to gain information on how semiclassical properties manifest themselves in the loop framework and how this could be exploited to do perturbation theory.

Since we do not know how to linearize LQG itself, our approach is indirect: we know how to linearize ADM gravity, and we know that field and loop-like degrees of freedom are physically related—after all they arise from a quantization of the same classical theory. We employ this relation to translate the free vacuum and free gravitons of ADM gravity into states of the loop representation. In contrast to earlier attempts in this direction[58, 59], we arrive at states in the Hilbert space of the *full non-linearized* theory. An important feature of the construction lies in the fact that it starts from momentum-regularized states and translates this property into a cutoff graph of the final loop states. Thus, they can be viewed as more or less coarse-grained states, depending on the value of the cutoff parameter.

Both the vacuum and gravitons take the form of Gaussian superpositions of spin networks whose graphs lie on the cutoff graph. We analyze the maximum of the Gaussian and find that spin networks at the peak have similar properties as weaves [62]: we determine their graphs and establish a relation between mean spin and cutoff scale. Our analysis indicates that in the limit of very small cutoff length, the peak spin networks become independent of the cutoff graph, have spin label $1/2$ and graphs of a length scale close to the Planck length. That is, the graphs of these spin networks maintain an effective Planck scale discreteness, while the mesh of the cutoff graph becomes infinitely fine.

The paper is organized as follows: in section 3.2, we linearize extended ADM gravity on a flat torus, quantize the reduced system and implement the linearized transformation to Ashtekar-Barbero variables. In this way, we obtain a free vacuum that is a functional of reduced triads or connections. Section 3.3 describes in several steps how we adapt this state to the degrees of freedom of loop quantum gravity. In section 3.4, we use the same procedure to define loop analogues of many-graviton states. Section 3.5 describes the peak analysis. In the final section, the construction and properties of the states are summarized, and we discuss the similarities and differences to other proposals for vacuum states. We also mention ideas on a genuine loop quantization of the free vacuum which would allow one to

remove the cutoff graph similarly as in the definition of Hamiltonian constraint and area operator. In the last subsection, we return to the issue of perturbation theory: we discuss how our states might be used for extracting a free part of the Hamiltonian constraint, and how they could be applied in renormalization. Interestingly, we arrive at an ansatz that is closely related to Smolin's string perturbations [70].

Notation and conventions

spatial indices: $a, b, c \ldots = 1, 2, 3$	LQG: loop quantum gravity
internal indices: $i, j, k \ldots = 1, 2, 3$	S: gauge-invariant spin network state
3-metric: g_{ab}	$\tilde{S}$: gauge-*variant* spin network state
determinant of 3-metric: g	$\mathcal{H}$: kinematic Hilbert space of LQG
Planck length: $l_p = \sqrt{\hbar\kappa}$	$\mathcal{H}_0$: gauge-inv. kinematic Hilbert space
gravitational coupling constant: $\kappa = 8\pi G/c^3$	$\mathcal{H}_{\text{diff}}$: gauge- and 3d-diff-invariant kinematic Hilbert space

We use units in which $c = 1$.

3.2 Vacuum of linearized Ashtekar-Barbero gravity

As we described in chapter 1, loop quantum gravity is based on a quantization of the Ashtekar-Barbero variables (A, E)—classical phase space variables that arise from a canonical transformation of the extended ADM variables (E, K) (see section 1.3.1 and 1.3.4). In order to arrive at some linearized version of LQG, we have, loosely speaking, six possibilities, corresponding to the different orders in which quantization (Q), linearization (L) and canonical transformation (C) can be applied. Let us abbreviate them by

1. LCQ 4. QLC
2. CLQ 5. QCL
3. LQC 6. CQL,

where the order of operations goes from left to right, starting with the classical ADM-theory and ending up with a linearized form of LQG. Combinations 4. and 5. are merely hypothetical, since the ADM-variables have never been rigorously quantized. Ideally, what we want is number 6., a linearization of full LQG. So far, however, we do not know how to do this (or, for that matter, what linearization should exactly mean in that case), since the quantum theory is formulated in terms of distributional fields or labelled networks. Hence, before that problem is resolved, we have to content ourselves with options 1. to 3. Actually, 1. is identical to 2. (what other meaning should be given to LC than CL?), so there remain possibilities CLQ and LQC: the former has been explored by Ashtekar, Rovelli and Smolin [58] for imaginary, and by Varadarajan [59] for real Immirzi parameter, when they applied a loop quantization to the linearized Ashtekar-Barbero variables. The route we follow in this paper takes the third variant LQC as its point of departure: we linearize the classical extended ADM-gravity, apply a Schrödinger quantization to it, and implement the linearized canonical transformation within the quantum theory. Thus, we arrive at a vacuum state which, at first glance, has little to do with loop-like degrees of freedom. The problem of relating this state to LQG will be the subject of section 3.3.

3.2.1 Linearization of classical extended ADM formulation

In classical theory, linearization rests, similarly as in quantum theory, on the idea that we choose a background, restrict attention to small deviations from it, and exploit this to give a lowest order approximation for the dynamics. In the following, we go through the classcial linearization of extended ADM gravity, but we will not motivate or derive each step. The procedure is similar to that for standard ADM and complex Ashtekar gravity, which has been described in the literature[2][74, 58].

To keep things as simple as possible, we choose space to be the 3-torus T^3 and linearize around a flat background on $T^3 \times \mathbb{R}$. The linearization consists of the following steps: we linearize the classical constraints, use them to obtain the reduced phase space, and determine the Poisson brackets and Hamiltonian on it. Once we have obtained the reduced classical system, we will quantize it using a Schrödinger representation (sec. 3.2.2) and, finally, apply the linearized form of the canonical transformation (sec. 3.2.3).

Recall that the extended ADM variables are the 1-density triad

$$E_i(\underline{x}) = E_i^a(\underline{x})\frac{\partial}{\partial x^a}\,, \quad i = 1,2,3, \tag{3.1}$$

and the canonically conjugate one-forms

$$K^i(\underline{x}) = K_a^i(\underline{x})\mathrm{d}x^a\,, \quad i = 1,2,3, \tag{3.2}$$

They are related to the 3-metric g_{ab} and extrinsic curvature K_{ab} by

$$gg^{ab} = E_i^a E_i^b\,, \tag{3.3}$$
$$K_a^k = K_{ab}E^{bk}/\sqrt{g}\,. \tag{3.4}$$

The Poisson brackets read

$$\left\{E_i^a(\underline{x}), E_k^b(\underline{y})\right\} = \left\{K_a^j(\underline{x}), K_b^k(\underline{y})\right\} = 0\,, \tag{3.5}$$

$$\left\{E_i^a(\underline{x}), K_b^k(\underline{y})\right\} = \frac{\kappa}{2}\,\delta_b^a\delta_i^k\,\delta(\underline{x}-\underline{y})\,. \tag{3.6}$$

The constraints are first-class and consist of the gauge, vector and Hamiltonian constraint:

$$G_{ij} = K_{a[i}\,E^a{}_{j]}\,, \tag{3.7}$$
$$V_a = \nabla_{[b}\left(K_{a]}{}^i\,E^b{}_i\right)\,, \tag{3.8}$$
$$C = -\frac{1}{\sqrt{E}}K_a{}^{[i}K_b{}^{j]}E^a{}_iE^b{}_j - \sqrt{E}\,R(E)\,. \tag{3.9}$$

E denotes the determinant $\det(E^a{}_i)$ and equals g. $R(E)$ stands for the 3d Riemann tensor when written as a function of the densitized triad.

Linearization around flat torus

Choose a flat classical background triad on T^3 such that the torus corresponds to a cube with macroscopic side length L and periodic boundary conditions. Moreover, choose, once and for all, a coordinate system in which the background E_i- and K^i-fields read

$$E^a_{cl\,k} = \delta^a_k\,, \quad K^k_{cl\,a} = K_{cl\,ab}E^{bk}_{cl}/\sqrt{g} = 0\,. \tag{3.10}$$

[2]For a detailed exposition of linearization in the Hamiltonian context, we refer the reader to [73].

We introduce the relative variables

$$e^a_k := E^a_k - \delta^a_k\,, \qquad K^k_a = K^k_a - 0$$

and adhere from now on to the convention that

$$e_{ka} \equiv e^a_k\,, \quad K^{ka} \equiv K^k_a\,. \tag{3.11}$$

That is, the spatial index of e can be freely moved between upper and lower right position, while the spatial index of K may be either in the lower or upper right position. The Poisson brackets of the relative variables are

$$\Big\{e_{ab}(\underline{x}), K^{cd}(\underline{y})\Big\} = \Big\{E^b_a(\underline{x}) - \delta^b_a, K^c_d(\underline{y})\Big\} = \frac{\kappa}{2}\,\delta^c_a\delta^d_b\,\delta(\underline{x}-\underline{y})\,. \tag{3.12}$$

By keeping only linear terms in e and K, we arrive at the linearized constraints

$$\begin{aligned} G^{ab} &= K^{[ab]}\,, && (3.13)\\ V^a &= -\partial_b K^{ba} + \partial^a K\,, && (3.14)\\ C &= 2\partial^a\partial^b e_{ab}\,, && (3.15) \end{aligned}$$

which are again first-class.

For the phase space reduction, it is convenient to change to Fourier space. On the 3-torus, we use the following conventions for Fourier series:

$$\begin{aligned} f(\underline{x}) &= \frac{1}{\sqrt{V}}\sum_{\underline{k}} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\, f(\underline{k}) \quad \text{and} && (3.16)\\ f(\underline{k}) &= \frac{1}{\sqrt{V}}\int \mathrm{d}^3x\; \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}}\, f(\underline{x})\,, && (3.17) \end{aligned}$$

where the wavevector $\underline{k}$ takes values in $2\pi/L\,\mathbb{Z}^3$. The delta functions on position and Fourier space have the respective transforms

$$\delta(\underline{x}-\underline{x}') = \frac{1}{V}\sum_{\underline{k}} \mathrm{e}^{\mathrm{i}\underline{k}\cdot(\underline{x}-\underline{x}')}\,. \tag{3.18}$$

and

$$\delta_{\underline{k},\underline{k}'} = \frac{1}{V}\int \mathrm{d}^3x\; \mathrm{e}^{-\mathrm{i}(\underline{k}-\underline{k}')\cdot\underline{x}}\,. \tag{3.19}$$

The Poisson bracket becomes

$$\Big\{e_{ab}(\underline{k}), K^{cd*}(\underline{k}')\Big\} = \frac{\kappa}{2}\,\delta^c_a\delta^d_b\,\delta_{\underline{k},\underline{k}'}\,. \tag{3.20}$$

By imposing the linearized constraints and choosing suitable gauge conditions, we require that e_{ab} and K^{ab} are symmetric, transverse and have a constant trace. This defines our reduced phase space. We denote the reduced variables by e^{red}_{ab} and K^{ab}_{red}, and write the Poisson bracket again as $\{\,,\,\}$.

When Fourier-transformed, the reduced variables can be decomposed into six zero-mode components, and two components for each nonzero $\underline{k}$, corresponding to the two polarizations of gravitational waves:

$$\begin{aligned} e^{\mathrm{red}}_{ab}(\underline{x}) &= \frac{1}{\sqrt{V}}\left(\sum_{i=1}^{6} e_i(0)\epsilon_{i\,ab}(0) + \sum_{k>0}\sum_{i=1}^{2} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\, e_i(\underline{k})\epsilon_{i\,ab}(\underline{k})\right)\,, && (3.21)\\ K^{ab}_{\mathrm{red}}(\underline{x}) &= \frac{1}{\sqrt{V}}\left(\sum_{i=1}^{6} K_i(0)\epsilon_{i\,ab}(0) + \sum_{k>0}\sum_{i=1}^{2} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\, K_i(\underline{k})\epsilon_{i\,ab}(\underline{k})\right)\,. && (3.22) \end{aligned}$$

k stands for the length of the vector $\underline{k}$. We specify the polarization tensors $\epsilon_{i\,ab}$ as follows: for each nonzero pair $\{\underline{k}, -\underline{k}\}$ we choose a right-handed coordinate system s.t. one of the vectors, say $\underline{k}$, points in the positive 3-direction. In this coordinate system, the polarization tensors are fixed as

$$\begin{aligned}
\epsilon_{1\,ab}(\underline{k}) &:= \frac{\mathrm{i}}{\sqrt{2}}\,(\delta_{1a}\delta_{2b} + \delta_{2a}\delta_{1b})\;, && (3.23)\\
\epsilon_{2\,ab}(\underline{k}) &:= \frac{1}{\sqrt{2}}\,(\delta_{1a}\delta_{1b} - \delta_{2a}\delta_{2b})\;, && (3.24)\\
\epsilon_{1\,ab}(-\underline{k}) &:= -\epsilon_{1\,ab}(\underline{k})\,, && (3.25)\\
\epsilon_{2\,ab}(-\underline{k}) &:= \epsilon_{2\,ab}(\underline{k})\,. && (3.26)
\end{aligned}$$

It follows that

$$\epsilon^*_{i\,ab}(\underline{k})\epsilon_{j\,ab}(\underline{k}) = \delta_{ij}\,, \tag{3.27}$$

and

$$\epsilon^*_{i\,ab}(\underline{k}) = \epsilon_{i\,ab}(-\underline{k})\,. \tag{3.28}$$

In the remainder, we only use the coordinate-independent properties (3.27) and (3.28), so apart from (3.23)–(3.26) all formulas apply to a general Cartesian coordinate system.

For $k = 0$, we take $\epsilon_{i\,ab}(0)$, $i = 1, \ldots, 6$, to be an orthonormal basis in the space of symmetric 2-tensors, i.e.

$$\epsilon^*_{i\,ab}(0)\epsilon_{j\,ab}(0) = \delta_{ij}\,. \tag{3.29}$$

The projection of an arbitrary two-tensor T_{ab} onto its reduced part is given by

$$(PT)_{ab}(\underline{x}) := \frac{1}{\sqrt{V}} \sum_{\underline{k}} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}} P_{ab}{}^{cd}(\underline{k}) T_{cd}(\underline{k})\,, \tag{3.30}$$

where

$$P_{ab}{}^{cd}(\underline{k}) = \epsilon_{i\,ab}(\underline{k})\epsilon_i^{cd*}(\underline{k})\,. \tag{3.31}$$

Recall that the Poisson brackets of the reduced phase space are the pull-back of the Poisson brackets on the full phase space. In our notation,

$$\Big\{e^{\mathrm{red}}_{ab}(\underline{x}), K^{cd}_{\mathrm{red}}(\underline{y})\Big\} = \Big\{(Pe)_{ab}(\underline{x}), (PK)^{cd}(\underline{y})\Big\}\,. \tag{3.32}$$

This implies that

$$\begin{aligned}
\Big\{e^{\mathrm{red}}_{ab}(\underline{k}), K^{cd*}_{\mathrm{red}}(\underline{k}')\Big\} &= \Big\{(Pe)_{ab}(\underline{k}), (PK)^{cd*}(\underline{k}')\Big\}\\
&= P_{ab}{}^{ef}(\underline{k}) P^{cd}{}_{gh}(\underline{k}') \Big\{e_{ef}(\underline{k}), K^{gh*}(\underline{k}')\Big\}\\
&= P_{ab}{}^{ef}(\underline{k}) P^{cd}{}_{gh}(\underline{k}')\, \frac{\kappa}{2}\delta^g_e\delta^h_f\delta_{\underline{k},\underline{k}'}\\
&= \frac{\kappa}{2} P_{ab}{}^{cd}(\underline{k})\delta_{\underline{k},\underline{k}'}\,. && (3.33)
\end{aligned}$$

Equivalently, we have

$$\Big\{e_i(\underline{k}), K^*_j(\underline{k}')\Big\} = \frac{\kappa}{2}\delta_{ij}\delta_{\underline{k},\underline{k}'} \tag{3.34}$$

for polarization and zero mode components.

Let us come to the linearized dynamics: the Hamiltonian is given by

$$H = \frac{1}{\kappa}\int \mathrm{d}^3x\; N_{cl}(\underline{x}) C_{\mathrm{quadr}}(\underline{x}) \tag{3.35}$$

where $C_{\rm quadr}$ is the quadratic part of the Hamiltonian constraint (3.9) when evaluated on the reduced phase space, and N_{cl} is the lapse density associated to the background: i.e. the lapse for which $\int \mathrm{d}^3x\ N(\underline{x})C(\underline{x})$ generates a flow that leaves the phase space point of the background fixed. In the case of the flat background, this is just $N_{cl} \equiv \sqrt{g}$. A straightforward calculation yields

$$H = \frac{1}{\kappa}\int \mathrm{d}^3x\ \left[K^{ab}_{\rm red}(\underline{x})K^{ab}_{\rm red}(\underline{x}) + \partial_c e^{\rm red}_{ab}(\underline{x})\partial_c e^{\rm red}_{ab}(\underline{x})\right]\ . \tag{3.36}$$

When expressed in terms of Fourier or polarization components, the Hamiltonian reads

$$\begin{aligned} H &= \frac{1}{\kappa}\sum_{\underline{k}}\left[K^{ab*}_{\rm red}(\underline{k})K^{ab}_{\rm red}(\underline{k}) + k^2\, e^{\rm red*}_{ab}(\underline{k})e^{\rm red}_{ab}(\underline{k})\right] &(3.37)\\ &= \frac{1}{\kappa}\sum_{\underline{k}}\left[K^*_i(\underline{k})K_i(\underline{k}) + k^2\, e^*_i(\underline{k})e_i(\underline{k})\right]\ . &(3.38)\end{aligned}$$

The polarization components for $\underline{k} \neq 0$ describe the spatial change in $e^{\rm red}_{ab}(\underline{x})$: they oscillate in harmonic potentials and always stay near $e_i(\underline{k}) = 0$. The zero modes are the constant part of the $e^{\rm red}_{ab}(\underline{x})$ field and move in a flat potential. This means that, to linear approximation, the overall shape of the torus behaves like a free particle. Unless the initial momentum is zero, the size of $e_i(0)$ will grow, so that at some point the linear approximation breaks down. This instability is due to the compactness of the torus. On $\mathbb{R}^3$, the zero modes are absent and the linearization stable.

3.2.2 Reduced phase space quantization

We quantize the reduced system by introducing operators $\hat{e}^{\rm red}_{ab}(\underline{k})$ and $\hat{K}^{ab}_{\rm red}(\underline{k})$, and replace the Poisson brackets (3.34) and (3.33) by commutation relations

$$\begin{aligned} \left[\hat{e}^{\rm red}_{ab}(\underline{k}), \hat{K}^{cd\dagger}_{\rm red}(\underline{k}')\right] &= \mathrm{i}\hbar\frac{\kappa}{2}P_{ab}{}^{cd}(\underline{k})\delta_{\underline{k},\underline{k}'}\ , &(3.39)\\ \left[\hat{e}_i(\underline{k}), \hat{K}^\dagger_j(\underline{k}')\right] &= \mathrm{i}\hbar\frac{\kappa}{2}\delta_{ij}\delta_{\underline{k},\underline{k}'}\,. &(3.40)\end{aligned}$$

The Hamilton operator becomes

$$\begin{aligned} \hat{H}_{\rm quadr} &= \frac{1}{\kappa}\sum_{\underline{k}}\left[\hat{K}^{ab\dagger}_{\rm red}(\underline{k})\hat{K}^{ab}_{\rm red}(\underline{k}) + k^2\,\hat{e}^{\rm red\dagger}_{ab}(\underline{k})\hat{e}^{\rm red}_{ab}(\underline{k})\right] &(3.41)\\ &= \frac{1}{\kappa}\sum_{\underline{k}}\left[\hat{K}^\dagger_i(\underline{k})\hat{K}_i(\underline{k}) + k^2\,\hat{e}^\dagger_i(\underline{k})\hat{e}_i(\underline{k})\right]\ . &(3.42)\end{aligned}$$

To represent these operators, we use a Schrödinger representation in terms of functionals of $e_i(\underline{k})$, where $\hat{e}_i(\underline{k})$ and $\hat{K}^i(\underline{k})$ act as multiplicative and derivative operators respectively:

$$\hat{e}_i(\underline{k}) \to e_i(\underline{k})\,, \quad \hat{K}_i(\underline{k}) \to -\mathrm{i}\hbar\frac{\kappa}{2}\frac{\partial}{\partial e^*_i(\underline{k})}\,. \tag{3.43}$$

Note that the coefficients $e_i(\underline{k})$ have to satisfy the reality condition $e^*_i(\underline{k}) = e_i(-\underline{k})$. That is, $e_i(0)$ is real, and for $\underline{k} \neq 0$ only half of the coefficients (say those for $k^1 > 0$) can be taken as independent variables. Thus, we choose the functional measure as

$$\int De := \left(\prod_{i=1}^{6}\int_{-\infty}^{\infty}\mathrm{d}e_i(0)\right)\left(\prod_{\underline{k}\neq 0,\, k^1>0}\ \prod_{i=1}^{2}\prod_{r=0}^{1}\int_{-\infty}^{\infty}\mathrm{d}e_{ir}(\underline{k})\right)\ . \tag{3.44}$$

The additional index $r = 0, 1$, denotes the real and imaginary part respectively.

Next we specify a free vacuum for the system. We require of it that it is time-independent and peaked around $e_i(\underline{k}) = 0$. The peakedness is necessary to ensure consistence with linearization[3]. At first, one might think that we look for the ground state of the Hamiltonian (3.42):

$$\Psi_G[e_i(\underline{k})] = \mathcal{N} \exp\left[-\frac{1}{\hbar\kappa}\sum_{\underline{k}} k\, e_i^*(\underline{k}) e_i(\underline{k})\right] . \tag{3.45}$$

For nonzero $\underline{k} \neq 0$, this functional is a Gaussian around $e_i(\underline{k}) = 0$ and satisfies our requirements. A dependence on the zero modes is missing, however, so Ψ_G has an infinite spread in $e_i(0)$, which is inconsistent with linearization. To remedy this, we add a Gaussian factor

$$\exp\left(-\frac{1}{\hbar\kappa}\,\omega_0\, e_i^2(0)\right) , \quad \omega_0 > 0\,, \tag{3.46}$$

for each zero mode. This gives us the new state

$$\Psi[e_i(\underline{k})] = \mathcal{N} \exp\left[-\frac{1}{\hbar\kappa}\sum_{\underline{k}} \omega(\underline{k})\, e_i^*(\underline{k}) e_i(\underline{k})\right] , \tag{3.47}$$

where

$$\omega(\underline{k}) = \begin{cases} k\,, & k > 0\,, \\ \omega_0\,, & k = 0\,. \end{cases} \tag{3.48}$$

Note that under evolution by $\hat{H}$, the spreading of the zero mode wavefunction proceeds on a time scale $\tau \sim 1/\omega_0$. By choosing ω_0 of the order $1/L$ or smaller, we can make the state Ψ practically time-independent for all microscopic processes.

In the following, we use this state as the free vacuum of ADM gravity on $T^3 \times \mathbb{R}$. (Throughout the text we write normalization factors unspecifically as $\mathcal{N}$ and do not keep track of their precise value.)

3.2.3 Canonical transformation

The classical Ashtekar-Barbero variables are obtained by the transformation

$$A_a^i = \beta K_a^i - \omega_a^i\,, \tag{3.49}$$

where β is the real Immirzi parameter and ω_a^i denotes the spin connection. As a function of E the latter reads

$$\omega_a^i[E] = -\frac{1}{2}\varepsilon^{ijk} E_k^b \left[\partial_b E_a^j - \partial_a E_b^j + E_j^c E_a^l \partial_b E_c^l\right] - \frac{1}{4}\varepsilon^{ijk}\left[2E_a^j \frac{\partial_b E}{E} - E_b^j \frac{\partial_a E}{E}\right] \tag{3.50}$$

The transformation leads to the new Poisson brackets

$$\left\{E_i^a(\underline{x}), A_b^j(\underline{y})\right\} = \left\{E_i^a(\underline{x}), \beta K_b^j(\underline{y}) - \omega_b^j[E](\underline{y})\right\} = \frac{\beta\kappa}{2}\,\delta_a^b \delta_i^j\, \delta(\underline{x} - \underline{y})\,, \tag{3.51}$$

or in terms of Fourier modes

$$\left\{E_i^a(\underline{k}), A_b^{j*}(\underline{k}')\right\} = \frac{\beta\kappa}{2}\,\delta_a^b \delta_i^j\, \delta_{\underline{k},\underline{k}'}\,. \tag{3.52}$$

[3]Of course, the peak property is just a minimum requirement: even then, the peak could be too wide, so that linearization is not applicable.

The linearization of (3.49) induces a canonical transformation on the reduced variables:

$$A^{ab}_{\rm red} = \varepsilon_{acd}\partial_c e^{\rm red}_{db} + \beta K^{ab}_{\rm red} \tag{3.53}$$

One may check that $A^{ab}_{\rm red}$ is again symmetric, transverse and of constant trace.

In the quantum theory, we introduce the new operator

$$\hat{A}^{ab}_{\rm red} = \varepsilon_{acd}\partial_c \hat{e}^{\rm red}_{db} + \beta \hat{K}^{ab}_{\rm red}\,. \tag{3.54}$$

Using that

$$\mathrm{i}\varepsilon_{acd}k_c\epsilon_{1\,db} = k\epsilon_{2\,ab} \quad \text{and} \quad \mathrm{i}\varepsilon_{acd}k_c\epsilon_{2\,db} = k\epsilon_{1\,ab}\,, \tag{3.55}$$

we see that (3.54) is equivalent to

$$\begin{array}{rcll} \hat{A}_1(\underline{k}) &=& k\hat{e}_2(\underline{k}) + \beta\hat{K}_1(\underline{k})\,, & \\ \hat{A}_2(\underline{k}) &=& k\hat{e}_1(\underline{k}) + \beta\hat{K}_2(\underline{k})\,, & \underline{k}\neq 0\,, \\ \hat{A}_i(0) &=& \beta\hat{K}_i(0)\,. & \end{array}$$

In the Schrödinger representation, $\hat{A}_i$ acts as

$$\begin{array}{rcll} \hat{A}_1(\underline{k}) &=& k\hat{e}_2(\underline{k}) - \mathrm{i}\hbar\dfrac{\kappa\beta}{2}\dfrac{\partial}{\partial e^*_1(\underline{k})}\,, & \\ \hat{A}_2(\underline{k}) &=& k\hat{e}_1(\underline{k}) - \mathrm{i}\hbar\dfrac{\kappa\beta}{2}\dfrac{\partial}{\partial e^*_2(\underline{k})}\,, & \underline{k}\neq 0\,, \\ \hat{A}_i(0) &=& -\mathrm{i}\hbar\dfrac{\kappa\beta}{2}\dfrac{\partial}{\partial e^*_i(0)}\,. & \end{array}$$

Up to an A-dependent phase (which we choose to be zero), eigenstates of $\hat{A}$ have the form

$$\varphi_A[e_i(\underline{k})] = \mathcal{N}\exp\left[\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{\underline{k}}\left(A^*_i(\underline{k})e_i(\underline{k}) - k\,e^*_1(\underline{k})e_2(\underline{k})\right)\right]\,. \tag{3.56}$$

Within the quantum theory, the canonical transformation (3.53) is implemented by a unitary map

$$\begin{array}{rcl} \psi[e] &\to& e^{\mathrm{i}f[e]/\hbar}\psi[e]\,, \\ \hat{O} &\to& e^{\mathrm{i}f[e]/\hbar}\hat{O}\,e^{-\mathrm{i}f[e]/\hbar}\,, \end{array}$$

that turns the $\hat{A}_i$-operator into a pure functional derivative in e^*_i, i.e.

$$\hat{A}_i(\underline{k}) \to -\mathrm{i}\hbar\frac{\kappa\beta}{2}\frac{\partial}{\partial e^*_i(\underline{k})}\,. \tag{3.57}$$

We see from (3.56) that the required factor is

$$e^{\mathrm{i}f[e]/\hbar} := \exp\left(\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{\underline{k}} k\,e^*_1(\underline{k})e_2(\underline{k})\right)\,.$$

The transformed vacuum state reads

$$\Psi[e_i(\underline{k})] = \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\sum_{\underline{k}}\left(\omega(\underline{k})\,e^*_i(\underline{k})e_i(\underline{k}) - \frac{2\mathrm{i}}{\beta}\,k\,e^*_1(\underline{k})e_2(\underline{k})\right)\right]\,, \tag{3.58}$$

In terms of reduced Fourier components, it takes the form

$$\Psi[e^{\rm red}_{ab}(\underline{k})] = \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\sum_{\underline{k}}\left(\omega(\underline{k})\, e^{\rm red*}_{ab}(\underline{k})e^{\rm red}_{ab}(\underline{k}) + \frac{1}{\beta}\, e^{\rm red*}_{ab}(\underline{k})\,\varepsilon_{acd}k_c e^{\rm red}_{db}(\underline{k})\right)\right]. \qquad (3.59)$$

By doing a Gaussian integration, we can transform (3.58) to the A-representation:

$$\Psi[A_i(\underline{k})] = \mathcal{N}\exp\left\{-\frac{1}{\hbar\kappa}\left[\frac{1}{\beta^2\omega_0}\, A_i^2(0) + \frac{3/4}{1+\beta^2}\sum_{k>0}\frac{1}{k}\left(A_i^*(\underline{k})A_i(\underline{k}) + \frac{2{\rm i}}{\beta}\, A_1^*(\underline{k})A_2(\underline{k})\right)\right]\right\}. \qquad (3.60)$$

The Schrödinger representation we have defined so far is formal, since arbitrarily high momenta appear. We can resolve this either by a rigorous continuum formulation (employing Gaussian measures on tempered distributions), or by using a regularization. Here, we choose the second possibility, as the existence of infinitely high momenta is physically questionable anyhow. We introduce an ultraviolet cutoff on the momenta, which we denote by Λ. We will sometimes refer to this as the regularized Fock representation, since it is the natural home for Fock-like excitations, as opposed to the polymer-like excitations in the loop representation. (Of course, strictly speaking we are dealing with a Schrödinger and not a Fock representation.)

With this adjustment, the canonically transformed and regularized vacuum becomes

$$\Psi[e^{\rm red}_{ab}(\underline{k})] = \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\sum_{k\leq\Lambda}\left(\omega(\underline{k})\, e^{\rm red*}_{ab}(\underline{k})e^{\rm red}_{ab}(\underline{k}) + \frac{1}{\beta}\, e^{\rm red*}_{ab}(\underline{k})\,\varepsilon_{acd}k_c e^{\rm red}_{db}(\underline{k})\right)\right]. \qquad (3.61)$$

In the regularized scheme, we consider the position space field $e^{\rm red}_{ab}(\underline{x})$ as a function

$$e^{\rm red}_{ab}(\underline{x}) := \frac{1}{\sqrt{V}}\sum_{k\leq\Lambda}{\rm e}^{{\rm i}\underline{k}\cdot\underline{x}}\, e^{\rm red}_{ab}(\underline{k})\,. \qquad (3.62)$$

of the Fourier modes $e^{\rm red}_{ab}(\underline{k})$. With that in mind, we can write the state functional also as

$$\begin{aligned}\Psi[e^{\rm red}_{ab}(\underline{k})] \;&=\; \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\left(\int{\rm d}^3x\int{\rm d}^3y\, W_\Lambda(\underline{x},\underline{y})\, e^{\rm red}_{ab}(\underline{x})e^{\rm red}_{ab}(\underline{y})\right.\right.\\ &\qquad\left.\left. -\;\frac{\rm i}{\beta}\int{\rm d}^3x\; e^{\rm red}_{ab}(\underline{x})\,\varepsilon_{acd}\partial_c e^{\rm red}_{db}(\underline{x})\right)\right], \qquad (3.63)\end{aligned}$$

where the kernel W_Λ is defined by

$$W_\Lambda(\underline{x},\underline{y}) = \frac{1}{V}\sum_{k\leq\Lambda}{\rm e}^{{\rm i}\underline{k}\cdot(\underline{x}-\underline{y})}\omega(\underline{k})\,. \qquad (3.64)$$

Let us repeat how we constructed this state: we linearized ADM gravity (L), quantized it (Q), determined the vacuum state and finally applied the canonical transformation (C). At the beginning of section 3.2, we referred to this as possibility "LQC". As far as we can see, the same state would be obtained if we performed the canonical transformation already at the classical level, and *then* applied the Schrödinger quantization. In that case, the determination of the vacuum functional is less transparent, because the Hamiltonian does not have the standard harmonic oscillator form.

Either way, the canonical transformation has the effect of adding a purely imaginary and β-dependent term in the exponential. The physical meaning of this term is not clear to us—to the same degree that we do not have a clear interpretation of the canonical transformation itself[4].

[4]There is a simple interpretation for transformations with *imaginary* parameter $\beta = \pm{\rm i}$. In that case, however, one is led to the Ashtekar formulation which involves complex connections.

3.3 Transition to degrees of freedom of LQG

At this point, we have a well-defined expression for a vacuum functional and we would like to translate it into a state of the LQG Hilbert space. In doing so, we want to preserve the physical properties of the Fock state as far as possible, which, includes, in particular, the cutoff on the momenta. The problem is that our state functional involves degrees of freedom that appear quite different from those of loop quantum gravity: on the one side, we have fields or their Fourier transforms, and on the other side abstract networks with spin labellings. In order to achieve a meaningful transition from Fock to loop state, it will be essential to find the right way to relate these degrees of freedom.

The Fock state was obtained by a reduced phase space quantization: we imposed the linearized gauge- and diff-constraint on the classical level, and then quantized the reduced degrees of freedom. Thus, we arrived at a state which is a functional of the reduced connection (see eqn. (3.60)).

LQG, on the other hand, is based on a quantization of the full phase space variables, yielding the kinematic Hilbert space $\mathcal{H}$. The full non-linear gauge- and diff-constraint are imposed subsequently to give the gauge- and diff-invariant Hilbert space $\mathcal{H}_0$ and $\mathcal{H}_{\text{diff}}$ respectively. The configuration space, from which the Schrödinger representation on $\mathcal{H}$ is built, consists of generalized connections— distributional connections with support on graphs, of which ordinary connections are only a special case. States in $\mathcal{H}_{\text{diff}}$ are functionals of gauge- and diff-equivalence classes of such generalized connections[5].

Logically, we can divide this difference between degrees of freedom into three steps:

1. reduced connection → connection,
2. connection → generalized connection,
3. generalized connection → gauge- and diff-equivalence class of generalized connections.

Our strategy for bridging this gap: We modify the Fock state Ψ such that it becomes a functional of connections (sec. 3.3.1 and 3.3.2); then we switch from a pure momentum regularization to a combined momentum / triangulation based regularization which gives us Ψ as a functional of generalized connections (sec. 3.3.3). In the final step, an averaging over the gauge- and diffeomorphism group has to be applied in order to arrive at a state in $\mathcal{H}_{\text{diff}}$. In this paper, we only do the gauge-averaging explicitly, which provides a state in $\mathcal{H}_0$ (sec. 3.3.4). The diff-projection remains to be done.

3.3.1 From reduced to full configuration space

We take the Fourier coefficients $E^a_i(\underline{k})$ as coordinates for the full configuration space. The position space field

$$E^a_i(\underline{x}) = \frac{1}{\sqrt{V}} \sum_{k \leq \Lambda} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\, E^a_i(\underline{k}) \tag{3.65}$$

should be understood as a function of the $E^a_i(\underline{k})$. Recall that e_{ia} without index $^{\text{red}}$ denotes the difference

$$e_{ia}(\underline{x}) \equiv e^a_i(\underline{x}) = E^a_i(\underline{x}) - \delta^a_i \tag{3.66}$$

[5]The standard procedure is to define states in $\mathcal{H}_0$ as gauge-invariant functionals of generalized connections, and states in $\mathcal{H}_{\text{diff}}$ are constructed as equivalence classes of states in $\mathcal{H}_0$ under diffeomorphisms. We obtain a mathematically equivalent formulation by defining everything in terms of equivalence classes of *connections*. Then, an element of $\mathcal{H}_{\text{diff}}$ is a functional of equivalence classes of connections w.r.t. gauge- and diff- transformations.

between the background triad and E^a_i. We introduce a Schrödinger representation for functionals on the full configuration space: the measure is defined by

$$\int DE := \prod_{k\leq\Lambda,\, k^1\geq 0} \prod_{i=1}^{3}\prod_{a=1}^{3}\prod_{r=0}^{1}\int_{-\infty}^{\infty} \mathrm{d}E^a_{ir}(\underline{k})\,, \tag{3.67}$$

and we represent the canonical commutation relations

$$\left[\hat{E}^a_i(\underline{k}), \hat{A}^{k\dagger}_b(\underline{k}')\right] = \mathrm{i}\hbar\frac{\beta\kappa}{2}\,\delta^b_a\delta^k_i\,\delta_{\underline{k},\underline{k}'}\,. \tag{3.68}$$

(cf. (3.52)) by setting

$$\hat{E}^a_i(\underline{k}) = E^a_i(\underline{k}) \quad \text{and} \quad \hat{A}^i_a(\underline{k}) = -\mathrm{i}\hbar\frac{\kappa\beta}{2}\,\frac{\partial}{\partial E^{a*}_i(\underline{k})}\,. \tag{3.69}$$

Analogously, there is an A-representation, where $\hat{E}^a_i(\underline{k})$ and $\hat{A}^i_a(\underline{k})$ are represented by

$$\hat{A}^i_a(\underline{k}) = A^i_a(\underline{k}) \quad \text{and} \quad \hat{E}^a_i(\underline{k}) = \mathrm{i}\hbar\frac{\kappa\beta}{2}\,\frac{\partial}{\partial A^{i*}_a(\underline{k})}\,. \tag{3.70}$$

To start with, we simplify the state (3.63) by dropping the β-dependent phase factor:

$$\Psi[e^{\mathrm{red}}_{ab}(\underline{k})] = \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x},\underline{y})\, e^{\mathrm{red}}_{ab}(\underline{x})e^{\mathrm{red}}_{ab}(\underline{y})\right]. \tag{3.71}$$

Thus, we avoid overly long formulas in the computations that follow. The treatment *with* phase factor is discussed in section 3.3.5.

Our aim is to extend the functional (3.71) to the full configuration space. The most simple possibility would be to use the projection map

$$e^{\mathrm{red}}_{ab}(\underline{k}) = P^{cd}_{ab}(\underline{k})e_{cd}(\underline{k})\,, \tag{3.72}$$

and define the extended state by the pull-back, i.e.

$$\Psi_{\mathrm{ext}}[E^a_i(\underline{k})] := \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x},\underline{y})\,(Pe)_{ab}(\underline{x})(Pe)_{ab}(\underline{y})\right]. \tag{3.73}$$

The problem with this state is that it has a very degenerate peak. In the vicinity of the background triad, this is ok because it corresponds to gauge-, diff- and time reparametrization invariance. If one goes farther away from the background triad, however, the linearized transformations are no longer symmetries of the theory. That means that if we follow long enough along the degenerate direction, we will arrive at triads that are very diff- and gauge-*in*equivalent to the chosen background, but they are still in the peak of the state functional due to the projector. That is a very unphysical property.

As an alternative, we could drop the projectors in (3.73) and define the state as

$$\Psi_{\mathrm{ext}}[E^a_i(\underline{k})] := \mathcal{N}\exp\left[-\frac{1}{\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x},\underline{y})\, e_{ab}(\underline{x})e_{ab}(\underline{y})\right]. \tag{3.74}$$

It implies that we make the Gaussian peak non-degenerate and throw out diff- and gauge-symmetry completely. After the state has been transferred to the LQG Hilbert space, the lost symmetries need to be restored. We can do this by applying an averaging over the

gauge and diffeomorphism group[6]. If we did the averaging in (3.74), it would effectively reintroduce the projection on the symmetric transverse part for small e, while for large e, it would establish the correct non-linear gauge- and diff-symmetry for the state. The problem: at the level of the loop state, gauge-averaging turns out to be rather complicated, if we start from (3.74).

In the present paper, we use a variation of this approach: to simplify the gauge-averaging, we replace the triad fluctuations $e_{ab}(\underline{x})$ in (3.74) by the fluctuation of the densitized inverse metric[7] $\tilde{g}^{ab}(\underline{x})$:

$$e_{ab}(\underline{x}) \quad \rightarrow \quad \frac{1}{2}\left(\tilde{g}^{ab}(\underline{x}) - \delta^{ab}\right) = \frac{1}{2}\left(E^a_i(\underline{x})E^b_i(\underline{x}) - \delta^{ab}\right) . \tag{3.75}$$

The triad fields are 1-densities, so $\tilde{g}^{ab}(\underline{x})$ has density weight 2. Since E^a_i contains only modes up to $k = \Lambda$, we can write $\tilde{g}^{ab}(\underline{x})$ also in a smeared form

$$\tilde{g}^{ab}_\Lambda := \int \mathrm{d}^3x' \; E^a_i(\underline{x})\delta_\Lambda(\underline{x} - \underline{x}')E^b_i(\underline{x}') \, , \tag{3.76}$$

where the smearing is done with the regularized delta function

$$\delta_\Lambda(\underline{x}) := \frac{1}{V}\sum_{k\leq\Lambda} \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}} \, . \tag{3.77}$$

The new state Ψ_{ext} is defined as

$$\begin{aligned}\Psi_{\mathrm{ext}}&[E^a_i(\underline{k})] \\ &:= \mathcal{N}\exp\left[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x \int \mathrm{d}^3y \; W_\Lambda(\underline{x},\underline{y}) \left(\int \mathrm{d}^3x' \; E^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')E^b_i(\underline{x}') - \delta^{ab}\right)\right. \\ &\qquad\left. \times \left(\int \mathrm{d}^3y' \; E^a_i(\underline{y})\delta_\Lambda(\underline{y}-\underline{y}')E^b_i(\underline{y}') - \delta^{ab}\right)\right] . \end{aligned} \tag{3.78}$$

Note that this state is almost gauge-invariant, but not completely, due to the smearing at the cutoff scale.

What is the justification for changing from the state (3.74) to (3.78)? With our substitution (3.75), the term

$$e_{ab}(\underline{x})e_{ab}(\underline{y}) \tag{3.79}$$

becomes

$$\frac{1}{2}\left(\int \mathrm{d}^3x' \; E^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')E^b_i(\underline{x}') - \delta^{ab}\right)\frac{1}{2}\left(\int \mathrm{d}^3y' \; E^a_k(\underline{y})\delta_\Lambda(\underline{y}-\underline{y}')E^b_k(\underline{y}') - \delta^{ab}\right) \tag{3.80}$$

$$\begin{aligned}
&= \frac{1}{4}\left(E^a_i(\underline{x})E^b_i(\underline{x}) - \delta^{ab}\right)\left(E^a_k(\underline{y})E^b_k(\underline{y}) - \delta^{ab}\right) \\
&= \frac{1}{4}\left[\left(\delta^a_i + e^a_i(\underline{x})\right)\left(\delta^b_i + e^b_i(\underline{x})\right) - \delta^{ab}\right]\left[\left(\delta^a_k + e^a_k(\underline{y})\right)\left(\delta^b_k + e^b_k(\underline{y})\right) - \delta^{ab}\right] \\
&= \frac{1}{4}\left[\delta^a_i\, e^b_i(\underline{x}) + e^a_i(\underline{x})\,\delta^b_i + e^a_i(\underline{x})\, e^b_i(\underline{x})\right]\left[\delta^a_k\, e^b_k(\underline{y}) + e^a_k(\underline{y})\,\delta^b_k + e^a_k(\underline{y})\, e^b_k(\underline{y})\right] \\
&= \frac{1}{4}\left(e^b_a(\underline{x}) + e^a_b(\underline{x})\right)\left(e^b_a(\underline{y}) + e^a_b(\underline{y})\right) + o(e^3) \\
&= e_{(ab)}(\underline{x})\, e_{(ab)}(\underline{y}) + o(e^3) \, .
\end{aligned} \tag{3.81}$$

[6]By dropping the projector in (3.74), we also lost the linearized scalar constraint: it is related to the dynamics and should reappear when the dynamics is linearized at the level of loops. For ideas in this direction, see section 3.6.4.

[7]We thank C. Rovelli for suggesting this modification.

The basic premise of linearization is that for a suitable coupling parameter, the values of e can be divided into "large" and "small" fluctuations with the following property: the "large" e are irrelevant because for these values the state functionals under consideration are exponentially damped. The remaining "small" fluctuations lie near the peak of the states and are small *enough* that to first approximation, higher orders in e can be neglected relative to the leading order terms. Thus, for small e, and within the precision of the linear approximation, (3.80) and (3.79) are equal except for the symmetrizers—a degeneracy which is due to the gauge-invariance in (3.78). Since we intend to apply a gauge-averaging anyhow, we can ignore this difference. For large e, both state functionals are exponentially damped, so that again the difference between (3.80) and (3.79) is not important.

3.3.2 "Complexifier" form

By a Gaussian integration, we transform the state (3.78) to the A-representation:

$$\begin{aligned}\Psi_{\text{ext}}[A^i_a(\underline{k})] \;=\; \mathcal{N} \int DE\; \exp\left(-\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{k\leq\Lambda} A^{a*}_i(\underline{k})E^a_i(\underline{k})\right)\\ \times\exp\Bigg[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x},\underline{y})\left(\int \mathrm{d}^3x'\; E^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')E^b_i(\underline{x}')-\delta^{ab}\right)\\ \times\left(\int \mathrm{d}^3y'\; E^a_i(\underline{y})\delta_\Lambda(\underline{y}-\underline{y}')E^b_i(\underline{y}')-\delta^{ab}\right)\Bigg]\,. \quad (3.82)\end{aligned}$$

Using that

$$\begin{aligned}&\int DE\; \exp\left(-\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{k\leq\Lambda} A^{a*}_i(\underline{k})E^a_{ir}(\underline{k})\right)\\ &= \left(\prod_{k\leq\Lambda,\,k^1>0}\prod_{i=1}^{3}\prod_{a=1}^{3}\prod_{r=0}^{1}\int_{-\infty}^{\infty}\mathrm{d}E^a_{ir}(\underline{k})\right)\exp\left(-\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{k\leq\Lambda}A^a_{ir}(\underline{k})E^a_{ir}(\underline{k})\right)\end{aligned}$$

is the delta functional on the connection, and that the operator $\hat{E}^{a\dagger}_i$ acts like

$$\mathrm{i}\hbar\frac{\kappa\beta}{2}\,\frac{\partial}{\partial A^i_a(\underline{k})}\,,$$

we can write the entire expression (3.82) as

$$\begin{aligned}\Psi_{\text{ext}}[A^i_a(\underline{k})] \;=\; \mathcal{N}\exp\Bigg[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x},\underline{y})\left(\int \mathrm{d}^3x'\; \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')\hat{E}^b_i(\underline{x}')-\delta^{ab}\right)\\ \times\left(\int \mathrm{d}^3y'\; \hat{E}^a_i(\underline{y})\delta_\Lambda(\underline{y}-\underline{y}')\hat{E}^b_i(\underline{y}')-\delta^{ab}\right)\Bigg]\,\delta(A)\,, \quad (3.83)\end{aligned}$$

where

$$\hat{E}^a_i(\underline{x}) = \frac{1}{\sqrt{V}}\sum_{k\leq\Lambda}\mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\,\hat{E}^a_i(\underline{k})\,. \quad (3.84)$$

This form of the state is similar, *but not identical*, to Thiemann's general complexifier form for coherent states [75]. Thiemann writes coherent states as

$$\exp\left(-\hat{C}\right)\,\delta(A-\tilde{A})\Big|_{\tilde{A}\to A^{\mathbb{C}}_{cl}} \quad (3.85)$$

where $A^{\mathbb{C}}_{cl}$ is the so-called complexified connection and contains the background triad. What we do here is somewhat different because we leave the background triad *outside* the delta-function. For that reason, it is a slight abuse of terminology if we call our analogue of $\hat{C}$, i.e.

$$\hat{C} := \frac{1}{4\hbar\kappa}\int \mathrm{d}^3x \int \mathrm{d}^3y\, W_\Lambda(\underline{x},\underline{y}) \left(\int \mathrm{d}^3x'\ \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')\hat{E}^b_i(\underline{x}') - \delta^{ab}\right) \times \left(\int \mathrm{d}^3y'\ \hat{E}^a_i(\underline{y})\delta_\Lambda(\underline{y}-\underline{y}')\hat{E}^b_i(\underline{y}') - \delta^{ab}\right), \quad (3.86)$$

again a complexifier.

3.3.3 From momentum cutoff to triangulation

At this point, we have a well-defined expression for a state functional of connection Fourier modes. In the next step, we turn this into a functional of generalized connections. We do so by a change of regularization scheme: similarly as one changes from a momentum to a lattice regularization in ordinary QFT, we trade the UV cutoff on the connection for a triangulation of space. Let $\mathcal{T}_\Lambda$ denote the simplicial complex of this triangulation. The connection is replaced by a generalized connection $\overline{A}$ on the dual complex[8] $\mathcal{T}^*_\Lambda$: a map that sends every edge e of $\mathcal{T}^*_\Lambda$ into a group element $g_e = \overline{A}(e)$. In making this transition, we want to alter the physical properties of the state as little as possible. The triangulation-based regularization should be such that it mimics the effects of the UV cutoff. For that reason, we choose $\mathcal{T}_\Lambda$ to be regular in the following sense:

When measured against the background metric δ_{ab}, the edges e of the dual complex should be straight and have lengths l_e in the range

$$(1-\epsilon)l_\Lambda \ \leq\ l_e \ \leq\ (1+\epsilon)l_\Lambda\,,$$

where

$$l_\Lambda := \frac{\pi}{\Lambda} \quad (3.87)$$

is the length scale corresponding to the cutoff Λ, and ϵ is some small fixed number.

The state we have so far consists of two parts: a delta functional of the Fourier coefficients of the connection and an operator acting on it. Let us first consider the delta functional: we replace it by the delta functional $\delta_{\mathcal{T}^*}(\overline{A})$ on $\mathcal{T}^*_\Lambda$, which is equal to a sum over all gauge-*variant* spin networks $\tilde{S}$ restricted to $\mathcal{T}^*_\Lambda$:

$$\delta_{\mathcal{T}^*}\left(\overline{A}\right) = \sum_{\tilde{S}\subset\mathcal{T}^*_\Lambda} \tilde{S}(0)\tilde{S}^*(\overline{A}) \quad (3.88)$$

In the operator part in (3.83), we have to replace the smeared operator product by an operator on functionals of $\overline{A}$, or equivalently, by an operator on spin network states. In other words: after having quantized

$$\tilde{g}^{ab}_\Lambda = \int \mathrm{d}^3x'\ E^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')E^b_i(\underline{x}') \quad (3.89)$$

in a Fock space manner, we will now quantize it along the lines of loop quantum gravity.

[8] Given $\mathcal{T}_\Lambda$, the dual complex is defined in the standard way, using the metric information of the background.

Loop quantization of smeared inverse densitized metric

There are several consistent ways in which one could quantize expression (3.89) on a triangulation. Of all the possibilities we will choose one that is very simple and gauge-symmetric.

Before dealing directly with (3.89), let us give a definition for $E^a_i(\underline{x})$. Spin network states are built from representation matrices $U_j(g_e)$ associated to edges e. To define $\hat{E}^a_i(\underline{x})$ on $U_j(g_e)$, we think of the latter as the path-ordered exponential of a connection $A^i_a(\underline{x})$, i.e.

$$\begin{aligned} U_j(g_e) &= \mathcal{P}\exp\left(\frac{\mathrm{i}}{\hbar}\int_e \mathrm{d}s\ \dot{e}^a(s)A^i_a(\underline{e}(s))\ \overset{(j)}{J}_i\right) \\ &= \sum_{n=0}^{\infty}\left(\frac{\mathrm{i}}{\hbar}\right)^n \int_0^1 \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \cdots \int_0^{s_{n-1}} \mathrm{d}s_n\ \dot{e}^{a_1}(s_1)A^{i_1}_{a_1}(\underline{e}(s_1))\cdots\dot{e}^{a_n}(s_n)A^{i_n}_{a_n}(\underline{e}(s_n))\ \overset{(j)}{J}_{i_1}\cdots\overset{(j)}{J}_{i_n}, \end{aligned}$$

and view the triad operator as the functional derivative $\mathrm{i}\hbar\kappa\beta/2\,\delta/\delta A^i_a(\underline{x})$. With this prescription, we obtain that

$$\hat{E}^a_i(\underline{x})U_j(g_e) = -\frac{\kappa\beta}{2}\int_0^1 \mathrm{d}s\ \dot{e}^a(s)\,\delta(\underline{x}-\underline{e}(s))\ U_j(g_{e_1(\underline{x})})\overset{(j)}{J}_i U_j(g_{e_2(\underline{x})}), \tag{3.90}$$

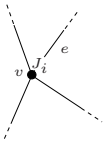
where $e_1(\underline{x})$ and $e_2(\underline{x})$ are the edges which result from splitting the edge e at the point $\underline{x}$. (When $\underline{x}$ is not on e, the definition of $e_1(\underline{x})$ and $e_2(\underline{x})$ is irrelevant, since the delta function gives zero.) The problem with (3.90) is that the original holonomy is split into two holonomies. Thus, when applied to a spin network on $\mathcal{T}^*_\Lambda$, the result will be a spin network which lives on the dual of a refined triangulation. If we want to stick to our original intention of defining the state on $\mathcal{T}^*_\Lambda$, we have to modify the action (3.90) such that it leaves the space of spin networks on $\mathcal{T}^*_\Lambda$ invariant.

The choice we take is

$$\hat{E}^a_i(\underline{x})U_j(g_e) = -\frac{\kappa\beta}{4}\left(\int_0^1 \mathrm{d}s\ \dot{e}^a(s)\right)\left(\delta(\underline{x}-\underline{e}(1))\ \overset{(j)}{J}_i\ U_j(g_e) + \delta(\underline{x}-\underline{e}(0))\ U_j(g_e)\overset{(j)}{J}_i\right). \tag{3.91}$$

This means that $\hat{E}^a_i$-operators can only create J's at the beginnings and ends of edges. In the obvious way, equation (3.91) generalizes to an action of $\hat{E}^a_i(\underline{x})$ on an entire spin network state on $\mathcal{T}^*_\Lambda$:

$$\hat{E}^a_i(\underline{x})\ \tilde{S} = \sum_v \sum_{\substack{\text{edges}\\ e\ \text{of}\ v}} F^a_{v,e}(\underline{x})$$



. (3.92)

The sum ranges over all vertices of the spin network, and the diagrams indicate where the J's are inserted. The *vertex-edge* form factor $F^a_{v,e}$ stands for

$$F^a_{v,e}(\underline{x}) := -\frac{\kappa\beta}{4}\left(\int_0^1 \mathrm{d}s\ \dot{e}^a(s)\right)\delta(\underline{x}-\underline{x}_v). \tag{3.93}$$

Given this definition of $\hat{E}^a_i(\underline{x})$, what would be a meaningful way to implement

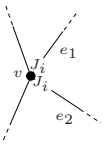
$$\hat{\tilde{g}}^{ab}_\Lambda \equiv \int \mathrm{d}^3x'\ \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')\hat{E}^b_i(\underline{x}') \quad ? \tag{3.94}$$

Roughly speaking, the smearing function $\delta_\Lambda(\underline{x}-\underline{x}')$ requires $\underline{x}$ and $\underline{x}'$ to be closer than l_Λ. Since the cutoff length l_Λ is also the length scale of the dual edges, we translate this into the condition that, if $\hat{E}^b_i(\underline{x}')$ inserts a J at a node v, $\hat{E}^a_i(\underline{x})$ can only insert J's at the same node.

This still allows for the possibility that $\hat{E}^a_i(\underline{x})$ and $\hat{E}^b_i(\underline{x}')$ insert J's on different edges of the same node. The action of the operator (3.94) becomes

$$\int \mathrm{d}^3x' \; \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')\hat{E}^b_i(\underline{x}') \; \tilde{S}$$

$$= \sum_v \sum_{\substack{\text{edges} \\ e_1,e_2 \text{ of } v}} \int \mathrm{d}^3x' \; F^a_{v,e_1}(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')F^a_{v,e_2}(\underline{x}') \quad . \tag{3.95}$$



Due to the contraction of i-indices, this action is gauge-invariant. For a generic basis of spin networks, it is not diagonal, however, and that makes its use complicated. In the following, we take a simpler choice, and adopt the point of view that $\hat{E}^a_i(\underline{x})\hat{E}^b_i(\underline{x}')$ should *not* involve contractions of J_i's from different edges. If we drop these cross-terms, the action of (3.94) is diagonal and simply reads

$$\int \mathrm{d}^3x' \; \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')\hat{E}^b_i(\underline{x}') \; \tilde{S} = \tilde{g}^{ab}_{\tilde{S}}(\underline{x}) \, \tilde{S} \, , \tag{3.96}$$

where $\tilde{g}^{ab}_{\tilde{S}}(\underline{x})$ is given by

$$\begin{aligned} \tilde{g}^{ab}_{\tilde{S}}(\underline{x}) &:= \sum_v \sum_{e \text{ of } v} \int \mathrm{d}^3x' \; F^a_{v,e}(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')F^b_{v,e}(\underline{x}') \, \hbar^2 j_e(j_e+1) \\ &\equiv \sum_v \sum_{e \text{ of } v} F^{ab}_{v,e}(\underline{x}) \, j_e(j_e+1) \, , \end{aligned} \tag{3.97}$$

and

$$\begin{aligned} F^{ab}_{v,e}(\underline{x}) &:= \int \mathrm{d}^3x' \; \hbar^2 F^a_{v,e}(\underline{x})\delta_\Lambda(\underline{x}-\underline{x}')F^b_{v,e}(\underline{x}') \\ &= \left(\frac{\hbar\kappa\beta}{4}\right)^2 \left(\int_0^1 \mathrm{d}s \; \dot{e}^a(s)\right)\left(\int_0^1 \mathrm{d}s' \; \dot{e}^b(s')\right) \int \mathrm{d}^3x' \; \delta(\underline{x}-\underline{x}_v) \, \delta_\Lambda(\underline{x}-\underline{x}') \, \delta(\underline{x}'-\underline{x}_v) \\ &= \left(\frac{l_p^2\beta}{4}\right)^2 \int_0^1 \mathrm{d}s \int_0^1 \mathrm{d}s' \;\; \dot{e}^a(s)\dot{e}^b(s')\delta_\Lambda(0) \, \delta(\underline{x}-\underline{x}_v) \, . \end{aligned} \tag{3.98}$$

We interpret $\tilde{g}^{ab}_{\tilde{S}}$ as a densitized inverse metric that is associated to the spin network $\tilde{S}$. This metric is distributional and has only support on vertices of the spin network graph. Thus, we can also write

$$\tilde{g}^{ab}_{\tilde{S}}(\underline{x}) = \sum_v \tilde{g}^{ab}_{\tilde{S}}(v) \, \delta(\underline{x}-\underline{x}_v) \tag{3.99}$$

where

$$\tilde{g}^{ab}_{\tilde{S}}(v) := \frac{\beta^2}{16} \, l_p^4 \delta_\Lambda(0) \sum_{e \text{ of } v} j_e(j_e+1) \int_0^1 \mathrm{d}s \int_0^1 \mathrm{d}s' \;\; \dot{e}^a(s)\dot{e}^b(s') \, . \tag{3.100}$$

With the definition (3.96) and the replacement of the delta functional by (3.88), we obtain the state

$$\begin{aligned} \Psi &= \mathcal{N} \sum_{\tilde{S} \subset \mathcal{T}^*_\Lambda} \tilde{S}(0) \exp\left[-\frac{1}{4\hbar\kappa} \int \mathrm{d}^3x \int \mathrm{d}^3y \; W_\Lambda(\underline{x}-\underline{y}) \left(\tilde{g}^{ab}_{\tilde{S}}(\underline{x}) - \delta^{ab}\right)\left(\tilde{g}^{ab}_{\tilde{S}}(\underline{y}) - \delta^{ab}\right)\right] \tilde{S}^* \\ &= \mathcal{N} \sum_{\tilde{S} \subset \mathcal{T}^*_\Lambda} \tilde{S}(0) \exp\left[-\frac{1}{4\hbar\kappa} \int \mathrm{d}^3x \int \mathrm{d}^3y \; W_\Lambda(\underline{x}-\underline{y}) \, \tilde{h}^{ab}_{\tilde{S}}(\underline{x}) \, \tilde{h}^{ab}_{\tilde{S}}(\underline{y})\right] \tilde{S}^* \, . \end{aligned} \tag{3.101}$$

In the second line, we abbreviated

$$\tilde{h}^{ab}_{\tilde{S}} := \tilde{g}^{ab}_{\tilde{S}} - \delta^{ab}\,. \tag{3.102}$$

Since spin networks on $\mathcal{T}^*_\Lambda$ are naturally mapped to spin networks in $\mathcal{H}$, the state (3.101) trivially extends to a state in $\mathcal{H}$. We denote this extended state by the same formula. Note that the restriction to spin networks on $\mathcal{T}^*_\Lambda$ in $\mathcal{H}$ is not ad hoc, but designed to preserve the momentum cutoff of the original Fock state. At the same time, the momentum regularization is not replaced completely, as it is still present in the kernel W_Λ and the factor $\delta_\Lambda(0)$.

3.3.4 Gauge projection

It remains to make the transition from (3.101) to a gauge-invariant state in $\mathcal{H}_0$. Gauge-averaging simply yields

$$\Psi_0 = \mathcal{N} \sum_{S\subset\mathcal{T}^*_\Lambda} S(0) \exp\left[-\frac{1}{4\hbar\kappa} \int \mathrm{d}^3x \int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\, \tilde{h}^{ab}_S(\underline{x})\, \tilde{h}^{ab}_S(\underline{y}) \right] S^*\,, \tag{3.103}$$

where the sum ranges over all gauge-*invariant* spin networks S on $\mathcal{T}^*_\Lambda$. We introduce the coefficient

$$\Psi_0(S) := S(0) \exp\left[-\frac{1}{4\hbar\kappa} \int \mathrm{d}^3x \int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\, \tilde{h}^{ab}_S(\underline{x})\, \tilde{h}^{ab}_S(\underline{y}) \right], \tag{3.104}$$

and write this more compactly as

$$\Psi_0 = \mathcal{N} \sum_{S\subset\mathcal{T}^*_\Lambda} \Psi_0(S) S^*\,. \tag{3.105}$$

One can think of $\Psi_0(S)$ as the wavefuntion of Ψ_0 in the S-representation.

When expressed in terms of Fourier coefficients of the two-tangent form factors $F^{ab}_{v,e}$, the coefficient reads

$$\Psi_0(S) = S(0) \exp\left[-\frac{1}{4l_p^2} \sum_{k\le\Lambda} \omega(\underline{k}) \left| \sum_v \sum_{e \text{ of } v} F^{ab}_{v,e}(\underline{k})\, j_e(j_e+1) - \sqrt{V}\delta^{ab}\delta_{\underline{k},0} \right|^2 \right]. \tag{3.106}$$

For Ψ_0 to be a well-defined state in the gauge-invariant Hilbert space $\mathcal{H}_0$, the norm

$$\|\Psi_0\|^2 = \mathcal{N}^2 \sum_{S\subset\mathcal{T}^*_\Lambda} |S(0)|^2 \exp\left[-\frac{1}{2l_p^2} \sum_{k\le\Lambda} \omega(\underline{k}) \left| \sum_v \sum_{e \text{ of } v} F^{ab}_{v,e}(\underline{k})\, j_e(j_e+1) - \sqrt{V}\delta^{ab}\delta_{\underline{k},0} \right|^2 \right] \tag{3.107}$$

has to be finite.

The number of possible spin network graphs on $\mathcal{T}^*_\Lambda$ is finite. Therefore, in order to show that (3.107) is finite, it suffices to prove that for every graph γ on $\mathcal{T}^*_\Lambda$,

$$\sum_{\substack{S \text{ with} \\ \text{graph } \gamma}} |S(0)|^2 \exp\left[-\frac{1}{2l_p^2} \sum_{k\le\Lambda} \omega(\underline{k}) \left| \sum_v \sum_{e \text{ of } v} F^{ab}_{v,e}(\underline{k})\, j_e(j_e+1) - \sqrt{V}\delta^{ab}\delta_{\underline{k},0} \right|^2 \right] < \infty\,. \tag{3.108}$$

Let us abbreviate $j_e(j_e+1)$ by c_e. The factor $S(0)$ gives a polynomial in the j_e's that depends on the connectivity of the graph γ. For a suitable polynomial $P[c_e]$ in c_e, we have

$$|S(0)|^2 \le P[c_e]\,. \tag{3.109}$$

Thus, an upper bound on (3.108) is given by

$$\left(\prod_e \sum_{j_e=1/2}^{\infty}\right) P[c_e] \exp\left(-\sum_{ee'} M_{ee'}\, c_e c_{e'} + \sum_e N_e c_e + K\right), \tag{3.110}$$

where $M_{ee'}$ is a symmetric positive matrix. By a linear transformation

$$\tilde{c}_e = T_{ee'} c_{e'}\,, \tag{3.111}$$

we render $M_{ee'}$ diagonal and arrive at the upper estimate

$$\left(\prod_e \sum_{\tilde{c}_e=0}^{\infty}\right) \tilde{P}[\tilde{c}_e] \exp\left(-\sum_e \lambda_e \tilde{c}_e^2 + \sum_e \tilde{N}_e \tilde{c}_e\right), \tag{3.112}$$

where the eigenvalues λ_e are positive and $\tilde{P}$ is a polynomial in $\tilde{c}_e$. By using further estimates with integrals, (3.112) can be shown to be convergent.

We conclude that the state (3.103) is an element in the gauge-invariant Hilbert space $\mathcal{H}_0$.

3.3.5 Inclusion of phase factor

Equation (3.103) is not yet the final form of the LQG state: what is missing is the contribution from the phase factor

$$\exp\left[\frac{\mathrm{i}}{\hbar\kappa\beta}\int \mathrm{d}^3x\; e^{\mathrm{red}}_{ab}(\underline{x})\,\varepsilon_{acd}\partial_c e^{\mathrm{red}}_{db}(\underline{x})\right] \tag{3.113}$$

that we have ignored so far in our transition from Fock to loop state. The treatment of the additional integral is quite analogous to that of the Gaussian term. There is only one difference—a difficulty that arises when implementing the differential operator ∂_c on the triangulation $\mathcal{T}^*_\Lambda$. If we were using a hypercubic lattice, we would simply get

$$\begin{aligned}\Psi_0(S) \;=\; & S(0)\exp\Bigg[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x \int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\,\tilde{h}^{ab}_S(\underline{x})\,\tilde{h}^{ab}_S(\underline{y}) \\ & + \frac{\mathrm{i}}{4\hbar\kappa\beta}\sum_v \tilde{h}^{ab}_S(v)\,\varepsilon^{acd}\,\nabla^c \tilde{h}^{ab}_S(v)\Bigg],\end{aligned} \tag{3.114}$$

where ∇^c stands for the lattice derivative in the c-direction. On a triangulation an implementation of ∂_c is more complicated and requires additional weighting factors to take account of the geometry of the triangulation. We do not determine this in detail and content ourselves with saying that the final state has the wavefunction

$$\Psi_0(S) = S(0)\exp\left[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x \int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\,\tilde{h}^{ab}_S(\underline{x})\,\tilde{h}^{ab}_S(\underline{y}) + \text{phase term}\right], \tag{3.115}$$

where the phase term is an analogue of

$$\frac{\mathrm{i}}{4\hbar\kappa\beta}\sum_v \tilde{h}^{ab}_S(v)\,\varepsilon^{acd}\,\nabla^c \tilde{h}^{ab}_S(v) \tag{3.116}$$

on a triangular lattice.

3.4 Graviton states

The steps that led us from the vacuum of linearized gravity to the state Ψ_0 can be repeated in complete analogy for gravitons. Let us go back to the Schrödinger representation of linearized extended ADM gravity (see 3.2.2). For $\underline{k} \neq 0$, we define creation and annihilation operators

$$\begin{aligned} a_i(\underline{k}) &:= \sqrt{\frac{k}{\hbar\kappa}}\, e_i(\underline{k}) + \frac{\mathrm{i}}{\sqrt{\hbar\kappa k}}\, K_i(\underline{k})\,, \\ a_i^\dagger(\underline{k}) &:= \sqrt{\frac{k}{\hbar\kappa}}\, e_i^\dagger(\underline{k}) - \frac{\mathrm{i}}{\sqrt{\hbar\kappa k}}\, K_i^\dagger(\underline{k})\,, \end{aligned}$$

such that

$$\Big[a_i(\underline{k}), a_j^\dagger(\underline{k}')\Big] = \delta_{ij}\delta_{\underline{k},\underline{k}'}\,. \tag{3.117}$$

The one-graviton state with polarization i and momentum $\underline{k}$ reads

$$\Psi_{i,\underline{k}}[e_i(\underline{k})] = a_i^\dagger(\underline{k})\Psi[e_i(\underline{k})] = 2\sqrt{\frac{k}{\hbar\kappa}}\, e_i^*(\underline{k})\, \Psi[e_i(\underline{k})] \tag{3.118}$$

As usual, we can write this also with tensors:

$$\Psi_{i,\underline{k}}[e^{\mathrm{red}}_{ab}(\underline{k})] = 2\sqrt{\frac{k}{\hbar\kappa}}\, \epsilon_i^{ab}(\underline{k}) e^{\mathrm{red}*}_{ab}(\underline{k})\, \Psi[e^{\mathrm{red}}_{ab}(\underline{k})] \tag{3.119}$$

The canonical transformation to Ashtekar-Barbero variables adds the phase factor

$$\exp\left(\frac{2\mathrm{i}}{\hbar\kappa\beta}\sum_{\underline{k}} k\, e_1^*(\underline{k}) e_2(\underline{k})\right)\,. \tag{3.120}$$

Next we extend the functional from the reduced to the full configuration space, as we did in sec. 3.3.1. This gives us

$$\begin{aligned} \Psi_{i,\underline{k}}[E^a_l(\underline{k})] &= 2\sqrt{\frac{k}{\hbar\kappa}}\, \epsilon_i^{la}(\underline{k})\, e^*_{la}(\underline{k})\, \Psi[E^a_j(\underline{k})] \\ &= 2\sqrt{\frac{k}{\hbar\kappa}}\, \epsilon_i^{la}(\underline{k})\, \frac{1}{\sqrt{V}} \int \mathrm{d}^3x\, \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}}\, e_{la}(\underline{x})\, \Psi[E^a_l(\underline{k})]\,. \end{aligned} \tag{3.121}$$

We replace $e_{la}(\underline{x})$ by $\frac{1}{2}\left(\tilde{g}^{la}_\Lambda(\underline{x}) - \delta^{la}\right)$ (using again the argument that higher orders in e can be neglected), and arrive at

$$\Psi_{i,\underline{k}}[E^a_l(\underline{k})] = \sqrt{\frac{k}{\hbar\kappa}}\, \epsilon_i^{la}(\underline{k})\, \frac{1}{\sqrt{V}} \int \mathrm{d}^3x\, \mathrm{e}^{\mathrm{i}\underline{k}\cdot\underline{x}} \left(\tilde{g}^{la}_\Lambda(\underline{x}) - \delta^{la}\right) \Psi[E^a_l(\underline{k})]\,. \tag{3.122}$$

We bring this into the complexifier form, make the transition to $\mathcal{H}$ and finally apply the gauge projector. The result is the state

$$\begin{aligned} \Psi_{i,\underline{k}} &= \mathcal{N} \sum_{S\subset\mathcal{T}^*_\Lambda} S(0) \sqrt{\frac{k}{\hbar\kappa}}\, \epsilon_i^{la}(\underline{k})\, \tilde{h}^{la*}_S(\underline{k}) \\ &\quad \times \exp\left[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x \int \mathrm{d}^3y\, W_\Lambda(\underline{x}-\underline{y})\, \tilde{h}^{ab}_S(\underline{x})\, \tilde{h}^{ab}_S(\underline{y}) + \text{phase term}\right] S^* \end{aligned} \tag{3.123}$$

in the gauge-invariant Hilbert space $\mathcal{H}_0$. We define an associated wavefunction

$$\begin{aligned}\Psi_{i,\underline{k}}(S) &:= S(0)\sqrt{\frac{k}{\hbar\kappa}}\,\epsilon_i^{la}(\underline{k})\,\tilde{h}_S^{la*}(\underline{k}) \\ &\quad\times\exp\Big[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\,\tilde{h}_S^{ab}(\underline{x})\,\tilde{h}_S^{ab}(\underline{y}) + \text{phase term}\Big]\,, \quad (3.124)\end{aligned}$$

and write the state as

$$\Psi_{i,\underline{k}} = \mathcal{N}\sum_{S\subset\mathcal{T}_\Lambda^*}\Psi_{i,\underline{k}}(S)S^*\,. \quad (3.125)$$

In the same way, we construct multiply excited states. Denote the polarizations and momenta of the gravitons by $i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N$. Then, the N-graviton state in $\mathcal{H}_0$ becomes

$$\Psi_{i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N} = \mathcal{N}_{i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N}\sum_{S\subset\mathcal{T}_\Lambda^*}\Psi_{i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N}(S)S^*\,, \quad (3.126)$$

where

$$\begin{aligned}\Psi_{i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N}(S) &:= S(0)\left(\prod_{n=1}^N\sqrt{\frac{k_n}{\hbar\kappa}}\,\epsilon_{i_n}^{la}(\underline{k}_n)\,\tilde{h}_S^{la*}(\underline{k}_n)\right) \\ &\quad\times\exp\Big[-\frac{1}{4\hbar\kappa}\int \mathrm{d}^3x\int \mathrm{d}^3y\; W_\Lambda(\underline{x}-\underline{y})\,\tilde{h}_S^{ab}(\underline{x})\,\tilde{h}_S^{ab}(\underline{y}) + \text{phase term}\Big]\,. \quad (3.127)\end{aligned}$$

The normalization factor $\mathcal{N}_{i_1,\underline{k}_1;\ldots;i_N,\underline{k}_N}$ depends on the excitation number of each mode.

3.5 Semiclassical properties of the vacuum state

By construction, the form of the vacuum Ψ_0 is similar to that of the Fock space functional we started from. The exponential is still of a Gaussian type, where now the role of the fluctuation variable is played by the spin networks S and their associated (inverse densitized) metric $\tilde{g}_S^{ab}$. As before, these fluctuations are non-locally correlated by the kernel W_Λ. It is immediate from (3.115) that for most spin networks the coefficient $\Psi_0(S)$ is exponentially damped. An absolute value of the order 0.1 to 1 is only attained for a relatively small class of spin networks: by analogy with quantum mechanics, we say that these spin networks constitute the "peak region" of Ψ_0. The "position" of the peak itself is given by those spin networks for which $\Psi_0(S)$ is exactly 1.

The state inherits its Gaussian property from the semiclassical peakedness of the original Fock state. This suggests that we interpret the peakedness of Ψ_0 as the way in which semiclassicality manifests itself on the level of spin networks: i.e. we interpret spin networks in the peak region as semiclassical fluctuations, and the spin networks at the peak position as the classical configuration.

>From that point of view, it would be interesting to know where exactly the peak is located; that is, for which spin networks S the coefficient $\Psi_0(S)$ reaches its maximum. Below we analyze this question and try to estimate the peak position: at first for general values of the cutoff length, and then, in section 3.5.2, for the limit where l_Λ is much smaller than the Planck length.

3.5.1 Peak position

We have three length scales: the Planck length l_p, the length cutoff $l_\Lambda = \pi/\Lambda$ corresponding to the momentum cutoff Λ, and the size of the 3-torus $L = V^{1/3}$. The Fourier coefficient of

the two-tangent form factor is

$$F^{ab}_{v,e}(\underline{k}) = \frac{1}{\sqrt{V}} \left(\frac{l_p^2 \beta}{4}\right)^2 \int_0^1 \mathrm{d}s \int_0^1 \mathrm{d}s' \; \dot{e}^a(s) \dot{e}^b(s') \delta_\Lambda(0) \, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} \,. \tag{3.128}$$

The factor $\delta_\Lambda(0)$ gives

$$\begin{aligned} \delta_\Lambda(0) &= \frac{1}{V} \sum_{k \le \Lambda} 1 \approx \frac{1}{V} \frac{1}{(2\pi/L)^3} \int\limits_{k \le \Lambda} \mathrm{d}^3 k \\ &= \frac{1}{(2\pi)^3} \frac{4}{3} \pi \Lambda^3 = \frac{\pi}{6} \frac{1}{l_\Lambda^3} \end{aligned} \tag{3.129}$$

In section 3.3.3, we assumed that the edges of the triangulation are straight and that their lengths are more or less equal to l_Λ. Hence we can approximate (3.128) by

$$\begin{aligned} F^{ab}_{v,e}(\underline{k}) &\approx L^{-3/2} \frac{\beta^2}{16} \, l_p^4 \, l_\Lambda n_e^a \, l_\Lambda n_e^b \, \frac{\pi}{6} \, \frac{1}{l_\Lambda^3} \, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} \\ &= \eta \, n_e^a n_e^b \, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} \,. \end{aligned} \tag{3.130}$$

Here, η stands for

$$\eta := \frac{\pi}{96} \, \beta^2 \, L^{-3} \, l_p^4 \, l_\Lambda^{-1} \tag{3.131}$$

and $\underline{n}_e$ denotes the normalized direction vector of the edge e:

$$\underline{n}_e := \frac{\underline{e}}{\sqrt{\underline{e}^2}} \tag{3.132}$$

The j_e-dependence of the $S(0)$-factor in (3.106) is polynomial. When determining the peak of the state, we can neglect it relative to the exponential dependence. Therefore, the peak condition becomes

$$\begin{aligned} &\sum_{k \le \Lambda} \omega(\underline{k}) \left| \eta \sum_v \sum_{e \text{ of } v} c_e \, n_e^a n_e^b \, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} - \delta^{ab} \, \delta_{\underline{k},0} \right|^2 \\ = \; & \omega_0 \left(\sum_v \sum_{e \text{ of } v} \eta \, c_e \, n_e^a n_e^b - \delta^{ab} \right)^2 + \sum_{0 < k \le \Lambda} \omega(\underline{k}) \left| \sum_v \sum_{e \text{ of } v} \eta \, c_e \, n_e^a n_e^b \, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} \right|^2 = \text{ minimal} \,. \end{aligned} \tag{3.133}$$

It is convenient to introduce the matrix

$$M^{ab} := \sum_v \sum_{e \text{ of } v} \eta \, c_e \, n_e^a n_e^b \,, \tag{3.134}$$

its traceless part

$$T^{ab} := M^{ab} - \frac{1}{3} \, M \, \delta^{ab} \,, \tag{3.135}$$

and the function

$$f^{ab}(\underline{x}) := \sqrt{V} \sum_v \sum_{e \text{ of } v} \eta \, c_e \, n_e^a n_e^b \, \delta(\underline{x} - \underline{x}_v) \,. \tag{3.136}$$

With this notation, condition (3.133) takes the form

$$\begin{aligned} &\omega_0 \left(M^{ab} - \delta^{ab} \right)^2 + \sum_{0 < k \le \Lambda} \omega(\underline{k}) \, |f^{ab}(\underline{k})|^2 \\ = \; & 3\omega_0 \left(\frac{1}{3} M - 1 \right)^2 + \omega_0 \, T^{ab} T^{ab} + \sum_{0 < k \le \Lambda} \omega(\underline{k}) \, |f^{ab}(\underline{k})|^2 = \text{ minimal} \,. \end{aligned} \tag{3.137}$$

If we minimize each term separately, we obtain the three conditions

$$\begin{aligned} &\text{(a)} \qquad T^{ab} = 0, \text{ i.e. } M^{ab} \text{ is isotropic.} \\ &\text{(b)} \qquad M = \sum_v \sum_{e \text{ of } v} \eta\, c_e = 3\,. \\ &\text{(c)} \qquad f^{ab}(\underline{k}) = 0 \text{ for } 0 < k \leq \Lambda. \end{aligned}$$

Condition (a) requires that on average the spin network edges are isotropic; (b) requires a certain mean value for $c_e = j_e(j_e+1)$, and (c) demands (weighted) homogeneity of spin network edges up to the scale l_Λ.

Let us first consider (b): we define the mean value of c_e on $\mathcal{T}^*_\Lambda$ by

$$\overline{c}_\Lambda = \frac{1}{N_e} \sum_e c_e\,, \tag{3.138}$$

where N_e is the total number of edges of the dual complex $\mathcal{T}^*_\Lambda$. Noting that N_e equals twice the total number N_v of vertices of $\mathcal{T}^*_\Lambda$, we can write condition (b) as

$$\overline{c}_\Lambda = \frac{3}{4\eta N_v}\,. \tag{3.139}$$

When a triangulation consists of regular tetrahedrons of side length a, the dual edge length a_* and the tetrahedron volume V_T are given by

$$a_* = \frac{1}{\sqrt{6}}\,a\,, \quad V_T = \frac{\sqrt{2}}{12}\,a^3 = \sqrt{3}\,a_*^3\,. \tag{3.140}$$

The tetrahedrons of our triangulation are very close to being regular, therefore

$$N_v \approx \frac{L^3}{\sqrt{3}\,l_\Lambda^3} \tag{3.141}$$

and

$$\begin{aligned} \overline{c}_\Lambda &= \frac{3}{4}\frac{96}{\pi}\,\beta^{-2}\,\frac{L^3\,l_\Lambda}{l_p^4}\,\frac{\sqrt{3}\,l_\Lambda^3}{L^3} \\ &\approx 40\,\beta^{-2}\left(\frac{l_\Lambda}{l_p}\right)^4. \end{aligned} \tag{3.142}$$

If we define the number j_Λ by

$$j_\Lambda(j_\Lambda+1) := 40\,\beta^{-2}\left(\frac{l_\Lambda}{l_p}\right)^4, \tag{3.143}$$

condition (b) becomes

$$\overline{c}_\Lambda = j_\Lambda(j_\Lambda+1)\,. \tag{3.144}$$

For certain values of l_Λ, the "spin" j_Λ is a half-integer (see Fig. 3.1). In these cases, it is easy to satisfy (a), (b) and (c): take the entire dual complex $\mathcal{T}^*_\Lambda$ as the spin network graph and label all edges with the spin j_Λ. Clearly, the mean value of c_e is $j_\Lambda(j_\Lambda+1)$, so (b) is fulfilled. Since we have chosen a very homogeneous and isotropic triangulation, the uniform spin distribution over all edges of $\mathcal{T}^*_\Lambda$ automatically implies isotropy (a) and homogeneity (c).

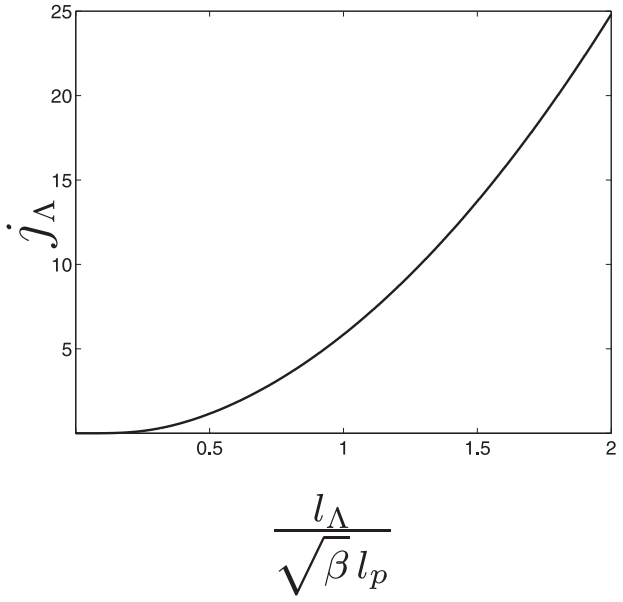


Figure 3.1: Dependence of j_Λ on cutoff length l_Λ.

Thus, for values of l_Λ which yield a half-integer j_Λ, the peak of the state (3.115) lies at spin networks which have the entire dual complex as graph and all spin labels equal to j_Λ. This peak exhibits a degeneracy due to the remaining freedom in choosing intertwiners at vertices. We see from (3.143) that the required spin depends strongly on the length cutoff l_Λ: beyond $l_\Lambda > 0.7\sqrt{\beta}\, l_p$ it grows quadratically in l_Λ. In the range $0.4\sqrt{\beta}\, l_p < l_\Lambda < 0.7\sqrt{\beta}\, l_p$, it takes values between $1/2$ and 10.

What are the peaks of the state when j_Λ lies between two half-integers $j_1 < j_2$? In that case, we have two opposing tendencies: condition (b) requires a mean value $\overline{c}_\Lambda = j_\Lambda(j_\Lambda + 1)$ between $j_1(j_1 + 1)$ and $j_2(j_2 + 1)$, and therefore at least an alternation of labels between j_1 and j_2. This is in conflict with condition (c), which demands spatial homogeneity on all scales larger or equal to l_Λ. That is, we cannot minimize all terms in (3.133) simultaneously.

There are essentially two possibilities how the peak state could behave: one is that it becomes some blend-over of the uniform spin distributions for $j_\Lambda = j_1$ and $j_\Lambda = j_2$. The other possibility: the configuration keeps the uniform spin j_1 while j_Λ moves away from j_1 towards j_2, and at some point it makes an abrupt transition to the uniform j_2-configuration.

3.5.2 Limit $l_\Lambda \ll l_p$

Instead of discussing this in detail for general l_Λ, we will concentrate on the most interesting case, namely, when l_Λ is much smaller than the Planck length. Then, j_Λ lies between 0 and $1/2$. To determine the peak states, we introduce three classes of spin networks that we call S_b, $S_{\not b}$ and $S_{ab\underline{c}}$.

S_b denotes the class of all spin networks on $\mathcal{T}^*_\Lambda$ that meet condition (b), whereas $S_{\not b}$ stands for those which violate (b). $S_{ab\underline{c}}$ consists of the spin networks which satisfy (a) and (b), and minimize the violation of condition (c), i.e. they minimize the third term in (3.137), while the first two terms are zero.

We will first determine the properties of spin networks in $S_{ab\underline{c}}$, and then argue that this class of states is the minimizing solution for condition (3.137).

We begin by analyzing condition (b). Suppose that for some $\Lambda' < \Lambda$, there is a triangulation of the type $\mathcal{T}_{\Lambda'}$ whose dual complex $\mathcal{T}^*_{\Lambda'}$ is coarser than $\mathcal{T}^*_\Lambda$ (for the definitons, see p.

76). Consider the restricted mean value

$$\overline{c}_{\Lambda'} := \frac{1}{N_{e'}} \sum_{e'} c_e \,, \tag{3.145}$$

where the sum runs only over edges of the coarser complex $\mathcal{T}^*_{\Lambda'}$, and $N_{e'}$ is the total number of its edges. By repeating the steps that led up to (3.144), we find that for spin networks on $\mathcal{T}^*_{\Lambda'}$, condition (b) is equivalent to

$$\overline{c}_{\Lambda'} = j_{\Lambda'}(j_{\Lambda'}+1) \,, \tag{3.146}$$

This equation generalizes equation (3.144) in the sense that it expresses (b) as a constraint on various scales of $\mathcal{T}^*_\Lambda$. In particular, for uniform spin distributions, it gives a relation between spin and scale of the distribution.

Using this reformulation of condition (b), we immediately find a large class of spin networks that solve it: according to (3.146), a spin network meets condition (b), if it has a dual complex $\mathcal{T}^*_{\Lambda'} \subset \mathcal{T}^*_\Lambda$ as its graph and a uniform spin labelling with $j_{\Lambda'}$. For given $j_{\Lambda'}$, we call the set of such spin networks $S_{j_{\Lambda'}}$. The admissible scales Λ' are those where $j_{\Lambda'}$ takes a half-integer value. For example, $j_{\Lambda'} = 1/2$ corresponds to the scale $0.4\sqrt{\beta}\, l_p$. By definition, a complex $\mathcal{T}^*_{\Lambda'}$ is nearly isotropic, so spin networks in $S_{j_{\Lambda'}}$ meet also condition (a).

The third condition on $S_{ab\underline{c}}$ states that the spin networks should minimize the term

$$\mathcal{I} = \sum_{0<k\leq\Lambda} \omega(\underline{k})\, |f(\underline{k})|^2 \,, \tag{3.147}$$

which measures inhomogeneity. A state in the class $S_{j_{\Lambda'}}$ breaks homogeneity (condition (c)) at the scale Λ'. Using (3.136) and (3.143), we see that it yields

$$\begin{aligned}
\mathcal{I} &= \sum_{0<k\leq\Lambda} \omega(\underline{k}) \left| \sum_v \sum_{e \text{ of } v} \eta\, c_e\, n_e^a n_e^b\, \mathrm{e}^{-\mathrm{i}\underline{k}\cdot\underline{x}_v} \right|^2 \sim \Lambda' \left[\eta\, j_{\Lambda'}(j_{\Lambda'}+1)\, N_{v'} \right]^2 \\
&\sim \Lambda' \left[\eta\, j_{\Lambda'}(j_{\Lambda'}+1) \left(\frac{l_\Lambda}{l_{\Lambda'}}\right)^3 N_v \right]^2 \\
&\sim \Lambda' \left[j_\Lambda^{-1}(j_\Lambda+1)^{-1}\, j_{\Lambda'}(j_{\Lambda'}+1) \left(\frac{l_\Lambda}{l_{\Lambda'}}\right)^3 \right]^2 \\
&\sim \Lambda' \left[\left(\frac{l_\Lambda}{l_p}\right)^{-4} \left(\frac{l_{\Lambda'}}{l_p}\right)^4 \left(\frac{l_\Lambda}{l_{\Lambda'}}\right)^3 \right]^2 \\
&= \Lambda' \left(\frac{l_{\Lambda'}}{l_\Lambda}\right)^2 \sim \frac{\Lambda^2}{\Lambda'} \,.
\end{aligned} \tag{3.148}$$

In going from the third to the fourth line, we used that $\eta N_v \sim j_\Lambda^{-1}(j_\Lambda+1)^{-1}$. Loosely speaking, high momentum wins over small momentum scale, and of all the sets $S_{j_{\Lambda'}}$ it is the one with the highest momentum scale Λ' that minimizes (3.147). Clearly, that set is $S_{1/2}$.

Given the domination of low spin in (3.148), it appears unlikely that configurations with mixed spin labels lead to the same or a lower value. If that expectation is correct, $S_{1/2}$ does not only minimize the sequence $S_{1/2}, S_1, S_{3/2}, \ldots$, but also all other states that satisfy (a) and (b). That is, $S_{1/2}$ equals $S_{ab\underline{c}}$. (When saying so, we are slightly imprecise: the minimizing conditions do not require that the graph takes *exactly* the form of a dual complex of a triangulation. The set $S_{ab\underline{c}}$ can also contain spin networks with other graphs $\mathcal{G}_{\Lambda'}$ as long as they are isotropic, homogeneous and have a length scale corresponding to $j_{\Lambda'} = 1/2$. When speaking of $S_{1/2}$ in the following, we mean to include these additional spin networks.)

Let us now consider the relation of $S_{ab\underline{c}}$ to the entirety of configurations on $\mathcal{T}^*_\Lambda$: we distinguish between configurations that satisfy (b) (the class S_b) and those which do not (the class $S_\not{b}$). By construction, $S_{ab\underline{c}}$ is the subset of S_b that comes closest to the minimum. Regarding $S_\not{b}$, there is, in principle, the possibility that the violation of (b) is compensated by greater homogeneity of the spin distribution. To check this, we compare expression (3.137) for $S_{ab\underline{c}}$ and $S_\not{b}$: for $S_{ab\underline{c}} = S_{1/2}$ we just get the inhomogeneity, i.e.

$$\mathcal{I} \sim \Lambda_p \left(\frac{l_p}{l_\Lambda}\right)^2 \sim \frac{l_p}{l_\Lambda^2}\,. \tag{3.149}$$

On the side of $S_\not{b}$, we obtain

$$\begin{aligned}
&3\omega_0 \left(\frac{1}{3}M - 1\right)^2 + \omega_0\, T^{ab}T^{ab} + \sum_{0<k\leq\Lambda} \omega(\underline{k})\,|f(\underline{k})|^2 \\
&\geq\ 3\omega_0 \left(\frac{1}{3}\sum_v \sum_{e\ \text{of}\ v} \eta\, c_e - 1\right)^2 \\
&\sim\ \omega_0 \left(\frac{4}{3}\,\eta\, N_v\, \overline{c}_\Lambda - 1\right)^2 \\
&\sim\ \omega_0\, \frac{1}{j_\Lambda^2(j_\Lambda+1)^2}\,[\overline{c}_\Lambda - j_\Lambda(j_\Lambda+1)]^2 \qquad (3.150)\\
&\sim\ \omega_0 \left(\frac{l_p}{l_\Lambda}\right)^8 [\overline{c}_\Lambda - j_\Lambda(j_\Lambda+1)]^2\,. \qquad (3.151)
\end{aligned}$$

If we think about states in $S_\not{b}$ whose mean value $\overline{c}_\Lambda$ has a finite difference to $j_\Lambda(j_\Lambda+1)$, the term (3.151) blows up much faster than (3.149) as $l_\Lambda \to 0$, so the $S_{1/2}$ states are preferred. Alternatively, we could consider states $S^r_\not{b}$ that have a fixed *relative* difference between $\overline{c}_\Lambda$ and $j_\Lambda(j_\Lambda+1)$, so that (3.150) remains constant. In that case, the inhomogeneity becomes the dominating criterion. Since $j_\Lambda(j_\Lambda+1)$ gets close to zero, $\overline{c}_\Lambda$ does so too, and the states $S^r_\not{b}$ become similar to S_b states. In fact, there is no obvious reason why such states should have advantage over the S_j-states. Hence we expect again that the $S_{1/2}$ class provides the lowest value.

This would mean that for $l_\Lambda \ll l_p$, the peak of the state consists of spin networks with the following properties: they are uniformly labelled by 1/2, their graphs lie on $\mathcal{T}^*_\Lambda$, are homogeneous and isotropic, and the effective length scale of the graphs is close to $\sqrt{\beta}\,l_p$, regardless of how fine $\mathcal{T}^*_\Lambda$ is.

3.6 Summary and discussion

3.6.1 Summary of results

The basic idea of our approach is to start from the free Fock vacuum of linearized gravity and construct from it a state Ψ_0 that could play the role of the "free" vacuum in loop quantum gravity. In making the transition from Fock to loop state, we have to take various choices that relate the field variables of the former to the polymer-like degrees of freedom of the latter. We have done so with the aim of making the state reasonably simple while preserving, as far as possible, the physical properties of the original state.

Let us repeat the logic of our construction:

1. We linearize extended ADM gravity around a flat background on $T^3 \times \mathbb{R}$.

2. We apply a reduced phase space quantization and specify the free vacuum.
3. We perform the canonical transformation to linearized Ashtekar-Barbero variables and implement it as a unitary transformation in the quantum theory.
4. We regularize the state by a momentum cutoff Λ.
5. The transition to the LQG Hilbert space is achieved in five steps:
   (a) We extend the state from a functional of reduced to a functional of full triads.
   (b) We replace the fluctuation $e_i^a = E_i^a - \delta_i^a$ by
   $$\frac{1}{2}\left(\tilde{g}^{ab}(\underline{x}) - \delta^{ab}\right) \equiv \frac{1}{2}\left(E_i^a(\underline{x})E_i^b(\underline{x}) - \delta^{ab}\right) ,$$
   the fluctuation in the densitized inverse metric associated to E_i^a.
   (c) We bring the state into the complexifier form.
   (d) We replace the momentum cutoff Λ by a regular triangulation $\mathcal{T}_\Lambda$ of length scale π/Λ, and thus obtain a functional of generalized connections.
   (e) Gauge projection yields a state in the gauge-invariant Hilbert space $\mathcal{H}_0$.

Ideally, this procedure should be completed by an averaging over the 3d-diff group, so that one receives a state in $\mathcal{H}_{\text{diff}}$. We have not carried out this step.

The state we get is a superposition

$$\Psi_0 = \mathcal{N} \sum_{S \subset \mathcal{T}_\Lambda^*} \Psi_0(S) S^* . \tag{3.152}$$

The sum ranges over all spin networks S whose graph lies on the dual complex $\mathcal{T}_\Lambda^*$ of the triangulation. The coefficients $\Psi_0(S)$ are given by

$$\Psi_0(S) = S(0) \exp\left[-\frac{1}{4\hbar\kappa} \int \mathrm{d}^3x \int \mathrm{d}^3y \; W_\Lambda(\underline{x}-\underline{y})\, \tilde{h}_S^{ab}(\underline{x})\, \tilde{h}_S^{ab}(\underline{y}) + \text{phase term} \right] . \tag{3.153}$$

To each spin network S we associate an inverse metric, and $\tilde{h}_S^{ab}$ stands for the difference between $\tilde{g}_S^{ab}$ and the flat background metric:

$$\tilde{h}_S^{ab} = \tilde{g}_S^{ab} - \delta^{ab} . \tag{3.154}$$

$\tilde{g}_S^{ab}$ is distributional and has only support on vertices of the spin network graph:

$$\tilde{g}_S^{ab}(\underline{x}) = \sum_v \tilde{g}_S^{ab}(v)\, \delta(\underline{x} - \underline{x}_v) \tag{3.155}$$

and

$$\tilde{g}_S^{ab}(v) = \frac{\beta^2}{16}\, l_p^4 \delta_\Lambda(0) \sum_{e \text{ of } v} j_e(j_e+1) \int_0^1 \mathrm{d}s \int_0^1 \mathrm{d}s' \;\; \dot{e}^a(s)\dot{e}^b(s') . \tag{3.156}$$

The functions $W_\Lambda(\underline{x}-\underline{y})$ and $\delta_\Lambda(\underline{x})$ are regularized forms of the kernel $1/|\underline{x}-\underline{y}|^4$ and the delta distribution respectively. On a hypercubic lattice, the phase term takes the form

$$\frac{\mathrm{i}}{4\hbar\kappa\beta} \sum_v \tilde{h}_S^{ab}(v)\, \varepsilon^{acd}\, \nabla^c \tilde{h}_S^{ab}(v) \tag{3.157}$$

where ∇^c denotes the lattice derivative. Here, we are working with triangulations, so this formula has to be adapted to the geometry of the triangular lattice. In complete analogy, we have also constructed N-graviton states.

In sec. 3.5, we investigated the S-dependence of the coefficient $\Psi_0(S)$: it is of Gaussian type and exponentially damped for most spin networks. At the peak of the Gaussian, the labelling is determined by a characteristic "spin" j_Λ which depends strongly on the cutoff scale (see Fig. 3.1):

$$j_\Lambda(j_\Lambda + 1) = 40\,\beta^{-2}\left(\frac{l_\Lambda}{l_p}\right)^4, \tag{3.158}$$

We have shown that for cutoff lengths l_Λ where j_Λ is a half-integer, the peak spin networks have the entire dual complex $\mathcal{T}^*_\Lambda$ as their graph and the edges are uniformly labelled by j_Λ. For values of l_Λ where j_Λ lies between two nonzero half-integers $j_1 < j_2$, the graph is again the entire complex and we expect that the spin labelling is either uniformly j_1, j_2, or a mixture of both. When l_Λ is much smaller than the Planck length, our analysis supports the idea that the peak spin networks become independent of the cutoff, have spin label $1/2$ and homogenous isotropic graphs at a length scale close to $\sqrt{\beta}\,l_p$.

3.6.2 Relation to other approaches

Let us point out some similarities and differences to what has already appeared in the literature:

The idea of using linearized states is not new and has been investigated before by Ashtekar, Rovelli & Smolin [58] and Varadarajan [59]. In both cases, however, the Fock states were transferred to a loop representation of *linearized* gravity, and not to the full representation, as we do here. Our state is also related to the class of complexifier coherent states which are used by Thiemann and collaborators [60]-[76]: what distinguishes our proposal is the particular choice of "complexifier" and the fact that we leave the background field outside the delta functional—a circumstance which facilitates the investigation of peak properties considerably. One could say that we keep the background in the "complexifier", though in the strict sense of the word this object is no complexifier anymore. Thiemann also introduces cutoff states that have support on a fixed graph, but it is not clear which graph should be chosen or how such a choice should be justified. In our case, the cutoff graph is fixed on physical grounds, namely, by the requirement that it emulates the momentum cutoff of the original Fock state.

Ashtekar and Lewandowski [61] have proposed a coherent state that is similar to ours in that it is based on J-insertions at vertices (see eqn. (3.95)). The way in which the J-operators (and form factors) are contracted is different, however. Moreover, their convolution kernel is $1/k$ (up to regulators), while ours corresponds to k. This difference can be explained as follows: Ashtekar and Lewandowski start out from $U(1)$ Maxwell theory, where the Hamiltonian is $H \sim E^2 + k^2B^2$, and then generalize from $U(1)$ to $SU(2)$. In the E-representation, the vacuum wavefunctional contains the kernel $1/k$ and that is also the kernel which appears in the complexifier. Our state, on the other hand, comes from the linearization of gravity where the linearized Hamiltonian reads $H \sim K^2 + k^2e^2$. Thus, the vacuum functional in e has k as its kernel and inherits it to the loop state.

Our results on the peak of the state clearly show a connection to the early weave approach [62]-[77], where a smeared metric operator was used to define eigenstates of the 3-metric. The construction of Ψ_0 involves a smeared inverse metric operator

$$\hat{\tilde{g}}^{ab}_\Lambda \equiv \int \mathrm{d}^3x'\ \hat{E}^a_i(\underline{x})\delta_\Lambda(\underline{x} - \underline{x}')\hat{E}^b_i(\underline{x}')\,, \tag{3.159}$$

and one of the conditions for maximizing $\Psi_0(S)$ consists in the requirement that S is an eigenstate of (3.159) with eigenvalue equal to δ^{ab}. Thus, the peak spin networks of Ψ_0 can be considered as a certain form of weave. For $l_\Lambda > \sqrt{\beta}\, l_p$, these weave states have the entire dual complex as graph, and the cutoff-dependence of the spin labelling follows the same logic as in the weave approach: for a large cutoff length l_Λ, the edges of the dual complex are dispersed far apart and they have to be labelled with large spins to generate a field strength equivalent to the background. For smaller l_Λ, the available edges are more densely distributed and a smaller mean spin suffices to attain the same field strength. It is somewhere near $l_\Lambda = \sqrt{\beta}\, l_p$ where the mean spin reaches values of the order 1.

If l_Λ is considerably smaller than the Planck length, one would expect that the average j_e has to be much smaller than the minimum nonzero spin $1/2$, so that only now and then an edge carries a spin $1/2$, while most other edges are labelled trivially. In other words, one could choose the triangulation increasingly finer than the Planck length, but the spin networks would stay apart as if they were trying to keep a triangulation fineness at a larger length scale. This is, in fact, what our analysis indicates: namely, that for $l_\Lambda \ll l_p$, the peak spin networks are in the fundamental representation and have isotropic homogeneous graphs whose edges maintain a distance scale close to the Planck scale. This stands in correspondence to the key result on weaves [62], which states that the approximation of 3-metrics cannot be improved by making the lattice finer than l_p. Let us mention that in [63], Corichi and Reyes have also constructed a state that is peaked around weave-like spin networks in the fundamental representation.

The form of our vacuum suggests that, in the spin network basis, the weaves at the peak and the spin networks near it play the role of the classical background and semiclassical fluctuations respectively. Thus, it would be interesting to investigate how this new notion of semiclassicality relates to the more standard definition in terms of mean values of operators and their fluctuations. Does our vacuum state produce mean values of area and volume that correspond to the classical background? Are the associated fluctuations small?

3.6.3 Continuum limit of free vacuum?

The cutoff-independence of the peak suggests that the entire state could become independent of l_Λ when l_Λ goes to zero. Note that the limit should be taken as

$$\lim_{\Lambda\to\infty} \mathcal{N}_\Lambda \sum_{S\subset\mathcal{T}^*_\Lambda} \Psi_0(S) S^* \tag{3.160}$$

where $\mathcal{N}_\Lambda$ normalizes the state for each value l_Λ in the sequence. Since $\tilde{g}^{ab}_S$ diverges for $l_\Lambda \to 0$, any deviation from the peak will be infinitely suppressed: the limiting state loses its spreading and becomes effectively a weave state.

The reason for this behaviour lies in the fact the we have not replaced the momentum cutoff completely when going from the Fock to the loop state. Instead we ended up with a mixed scheme where part of the regularization is provided by the cutoff graph $\mathcal{T}^*_\Lambda$ and part of it by the momentum cutoff in the kernel W_Λ and the factor $\delta_\Lambda(0)$ in (3.156). Such a mixed regularization may be ok for $l_\Lambda > l_p$, but it becomes physically questionable when the regulator is removed: then, the triangulation part leads to an effective cutoff at the Planck scale ("space is discrete"), while the momentum part lets $\delta_\Lambda(0)$ diverge ("space is continuous").

To arrive at a more meaningful continuum limit, it would be desirable to express the regularization entirely in terms of the triangulation. Carlo Rovelli has suggested that one might achieve this by starting from a modification of the Fock state (3.83): observe that the kernel W_Λ and the delta function δ_Λ require the background triad E_{cl} as an input, so

we write them as $W_{\Lambda,E_{cl}}$ and $\delta_{\Lambda,E_{cl}}$. In the state functional, we replace these quantities by $W_{\Lambda,E}$ and $\delta_{\Lambda,E}$, since it does not change the part of the exponential which is quadratic in the fluctuation $E - E_{cl}$. The kernel $W_{\Lambda,E}$ is the same as $\sqrt{-\Delta_{\Lambda,E}}$ where $\Delta_{\Lambda,E}$ denotes the Λ-regularized Laplace operator associated to the triad E.

The next step is the transition to the LQG state, that is, one has to decide on how to implement the propagator $W_{\Lambda,E} = \sqrt{-\Delta_{\Lambda,E}}$ in the LQG complexifier. The first thing to note is that when making this transition, the Λ in $\Delta_{\Lambda,E}$ can be dropped because the triangulation takes care of the momentum regularization. Therefore, what remains to be done is to find a loop quantization $\Delta_{\hat{E}}$ of Δ_E on this triangulation. If things are simple, it will act diagonally on spin networks and contributes a factor $W_{e,e'}(S)$ for every pair e, e' of edges of the spin network S on which the complexifier acts. In a similar way, one might be able to define a loop analogue of $\delta_\Lambda(0)$.

The conjecture is that by this procedure one arrives at a "free" vacuum that becomes independent of the cutoff when l_Λ goes to zero and *keeps* the form that it had for $l_\Lambda \sim l_p$.

3.6.4 Free theory, perturbation theory and renormalization?

We finally come to the big question that motivates this work: can one obtain a semiclassical perturbation series for loop quantum gravity, and if yes, how?

To start with, let us recall how perturbation theory works in ordinary QFT, or even simpler, in quantum mechanics. Consider the anharmonic oscillator with the Hamiltonian

$$H = \frac{1}{2}P^2 + \frac{1}{2}\omega^2(Q - Q_0)^2 + \lambda(Q - Q_0)^4 \,. \tag{3.161}$$

Assume that we want to do perturbation theory around the static solution $Q(t) = Q_0$, $P(t) = 0$. For that purpose, we introduce the relative variables

$$q := Q - Q_0 \,, \quad p := P \,, \tag{3.162}$$

and "expand" the Hamiltonian in q and p:

$$H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2 + \lambda q^4 \,. \tag{3.163}$$

Perturbation theory rests on the idea that we consider only processes that involve a certain subset of states; namely those whose wavefunction is peaked near the classical position and momentum, and strongly damped farther away from it. If the width of these states is sufficiently small, one may ignore higher order terms to first approximation, and use instead the free Hamiltonian

$$H_0 = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2 \,. \tag{3.164}$$

The associated free vacuum is

$$\psi_0(q) = \mathcal{N} \exp\left(-\frac{\omega}{2\hbar}q^2\right) \,. \tag{3.165}$$

For this approximation to be consistent, the width of Ψ_0 should be small enough that the neglection of higher orders is admissible. We check this by applying the full potential $\hat{V}$ to ψ_0:

$$\hat{V}\psi_0(q) = \mathcal{N}\left(\frac{1}{2}\omega^2 q^2 + \lambda q^4\right) \exp\left(-\frac{\omega}{2\hbar}q^2\right) \,. \tag{3.166}$$

The width of ψ_0 is $\sqrt{\hbar/\omega}$, so the consistency condition becomes

$$\frac{1}{2}\omega^2 \frac{\hbar}{\omega} \gg \lambda \left(\frac{\hbar}{\omega}\right)^2 \,, \tag{3.167}$$

or

$$\lambda \ll \frac{\omega^3}{2\hbar}\,. \tag{3.168}$$

If it is satisfied, there is hope that a perturbative treatment leads to meaningful results. Otherwise we are in the non-perturbative regime.

Let us transfer this logic to the linearization of ADM gravity on $\mathcal{T}^3 \times \mathbb{R}$. When we keep all orders in the reduced fluctuation variables, the Hamiltonian reads

$$H = \frac{1}{\kappa}\int \mathrm{d}^3x \left[\left(K_a^l K_b^j - K_a^j K_b^l\right)\left(\delta_j^a + e_j^a\right)\left(\delta_l^b + e_l^b\right) - \left|\det\left(\delta_j^a + e_j^a\right)\right| R\left(\delta_j^a + e_j^a\right)\right]\,. \tag{3.169}$$

Suppose we quantize this system as we did in the free case, i.e. by a standard canonical quantization and using a momentum cutoff Λ as a regulator. Then, a calculation analogous to (3.166) yields the expansion parameter

$$\alpha := \frac{\sqrt{\hbar\kappa_\Lambda}\, L^{3/2}}{l_\Lambda^{5/2}}\,. \tag{3.170}$$

Here, we have taken κ_Λ to be the gravitational coupling at the cutoff scale. L is the size of the torus and the time scale of the process we consider. The simplest way to relate κ_Λ to the classical value is dimensional scaling, i.e.

$$\kappa_\Lambda = \frac{l_\Lambda^2}{L^2}\,\kappa\,. \tag{3.171}$$

In that case, the parameter (3.170) becomes

$$\alpha = \frac{l_p\, L^{1/2}}{l_\Lambda^{3/2}}\,. \tag{3.172}$$

This estimate indicates that for L of the order 1cm, perturbation theory requires a cutoff length larger than 10^{-22}cm. For $l_\Lambda = l_p$, the theory would be highly non-perturbative.

Of course, what we actually want to analyze is loop quantum gravity, and not a conventional quantization of gravity with an arbitrary cutoff. With that aim in mind, we transferred the free vacuum of (3.169) to the LQG Hilbert space. The semiclassical spreading around $e = 0$ turned into a Gaussian spreading around weave-like spin networks. The more the spin labelling of a spin network differs from that of the weave, the more it is damped in the sum (3.152). The quantity that measures the deviation from the peak is

$$\tilde{h}_S^{ab} = \tilde{g}_S^{ab} - \delta^{ab}\,, \tag{3.173}$$

and stands in obvious correspondence to the fluctuation e of the traditional approach. Therefore, it appears natural to ask the following questions: could this correspondence give us a hint on how semiclassicality manifests itself in full loop quantum gravity? Could we take the size of $\tilde{h}_S^{ab}$, or a similar quantity, as the loop analogue of the fluctuation, and use it as a measure for telling whether a spin network is a "small" or a "large" fluctuation w.r.t. a semiclassical state? Suppose that we let the full Hamiltonian constraint $\hat{C}$ act on a state like Ψ_0 that is peaked around weaves. Would there be parts of $\hat{C}$ that are more relevant than others, in the same way that $\omega^2 q^2/2$ is more relevant than λq^4 when acting with the potential on $\exp(-\omega q^2/2\hbar)$? In other words, can we write $\hat{C}$ as a sum

$$\hat{C} = \hat{C}_0 + \hat{C}_1 \tag{3.174}$$

such that to first approximation $\hat{C}_1\Psi_0$ can be neglected relative to $\hat{C}_0\Psi_0$? To answer that question, one would have to analyze if the quantity

$$\Psi_0(S)\,\hat{C}S \tag{3.175}$$

can be approximated for spin networks close to the peak weaves. I.e. for those spin networks whose spin assignments differ from the weaves on only few edges and only by little spin.

Thus, we are led to an ansatz that appears very similar to the one used by Smolin when he analyzes string perturbations of causal spin networks [70]: in that case, the "differing" edges are chosen to form loops and are specified on an entire history of causal spin networks. The sequence of loop-like deviations is viewed as a string worldsheet, and the resulting variation in the amplitude is to some extent evaluated.

The inherent cutoff-property of LQG suggests that loop quantization is related to conventional quantization schemes where a cutoff $l_\Lambda = l_p$ is put in by hand. Similarly as there, the expansion parameter could be large and perturbation theory impossible. In that case, a coarse-graining procedure [66, 67] may be necessary to compute an effective Hamiltonian constraint for a lower scale l_Λ, where the expansion parameter is small. Our state Ψ_0 might be useful in such attempts because it is an element in the full Hilbert space and represents at the same time a semiclassical state at the cutoff scale l_Λ. One could try to extract an effective matrix element

$$\langle\Psi_0|P_\Lambda|\Psi_0\rangle \quad \text{from} \quad \langle\Psi_0|P|\Psi_0\rangle\,, \tag{3.176}$$

where P is the "bare" projector onto the physical Hilbert space, and P_Λ is based on a renormalized Hamiltonian constraint at the cutoff scale $l_\Lambda < l_p$.

Admittedly, the considerations of this last section are vague and need to be concretized in many ways. We hope to do so in future work.

Chapter 4

Geometric spin foams, Yang-Mills theory and background-independent models

4.1 Introduction

The concept of spin foams is both new and old. It is old in the sense that it is just another name for the plaquette diagrams that appear in the strong coupling expansion of lattice gauge theories—a method that has been developed in the seventies (see e.g. [1]). It is also new, however, because it was reinvented in the nineties within the context of quantum gravity, and used to construct rigorous proposals for sums over geometries [78, 79].

In both cases, the sum over spin foam diagrams arises from character expansions on plaquettes, and a subsequent integration over group variables. Although this is a well-known procedure in the strong coupling expansion, it is less known that it can be also used to give an alternative *non-perturbative* definition of the original theory. For pure lattice gauge theory, this exact transformation from path integral to spin foam sum has been described by Oeckl & Pfeiffer [80].

In the first part of the article, we explain this *dual* transformation in a pedagogic and step-by-step fashion. Our presentation differs from that in [80] in that we lay emphasis on the fact that plaquettes can be organized into larger surfaces with a single representation label[1]. The different surfaces meet along lines, so the entire diagram takes the form of a branched surface with labels on its unbranched components. Thus, we are led to a more geometric definition of spin foams, where the spin foam is not identified with the lattice, but instead regarded as a branched surface that lies on it. Based on this geometric viewpoint, we obtain a particularly simple description of a spin foam model that is dual to lattice Yang-Mills theory in dimension $d \geq 2$: each unbranched component is weighted with $\exp(-T_\rho \mathcal{A})$, where $\mathcal{A}$ is the area of the surface and T_ρ a representation-dependent tension.

The second part of the paper concerns the spin foam approach to gravity: i.e. the attempt to employ spin foam sums for defining non-perturbative and background-independent models of quantum gravity. The construction of such models is usually plagued by the problem that it depends on the choice of a lattice or triangulation. The spin foam sum is restricted to spin foams that are congruent with the lattice. Unless the theory is topologically invariant, the latter stands in obvious contradiction to background and cutoff independence.

We propose a solution to this difficulty which is, again, based on the geometric notion of spin foams. There are spin foam models whose weight factors are independent of how spin foams are embedded on the lattice. Their amplitudes depend only on topological properties of the branched surface, but not on how its components and branching lines are subdivided by the lattice. We elevate this property to a symmetry principle, which appears natural from

[1] This is known in lattice field theory [1] and has been pointed out in the spin foam literature [81, 82].

the relational point of view: what counts is how different surfaces connect to each other, while subdivisions within a surface are physically irrelevant. When a spin foam model satisfies the symmetry requirement, we discard the lattice and extend the model to a sum over *all* branched and labelled surfaces in the manifold. The new sum is lattice-independent, but contains an infinite overcounting of homeomorphically equivalent configurations. We factor this gauge volume off, and arrive at a sum over *abstract* spin foams—equivalence classes of spin foams under homeomorphisms—which carry only topological and combinatorial information. As an example, we consider a version of the Barrett-Crane model, for which we check the symmetry property.

With this method, we propose an alternative to the group field theory approach [83, 84], where background-independence is achieved by a sum over lattices. Related ideas on gravity spin foams have been expressed by Bojowald & Perez [85] and Zapata [86]. Our abstract spin foams are similar to the combinatorial spin foams of the causal histories approach [71, 72]. The dual Yang-Mills model should be compared with attempts to formulate lattice Yang-Mills theory in terms of strings (see e.g. [87]).

The paper is organized as follows: in the next section, we set our framework and conventions for pure lattice gauge theory. In section 4.3, we recall basic properties of spin network states. The latter will be extensively used in section 4.4 where we explain the dual transformation. It is also there that the geometric definition of spin foams and its motivation will become clear. After that we present the spin foam model of Yang-Mills theory (sec. 4.5). In section 4.6, we consider spin foam models of gravity: we introduce the symmetry condition and describe the extension to the lattice-independent formulation. We verify in the appendix that it is applicable to one of the versions of the Barrett-Crane model. Section 4.7 contains the summary and discussion.

4.2 Lattice gauge theory

Consider a hypercubic lattice κ in $\mathbb{R}^d$ that has lattice constant a and finite side lengths L_i, $i = 1, \ldots, n$. The dimension d should be greater than 1. Let us choose an orientation for each link and plaquette of the lattice, and call the resulting oriented links and plaquettes *edges* e and *faces* f respectively. The choice of this orientation is arbitrary and all physical quantities are independent of it. We write E_κ for the set of all edges of κ. The edges on the boundary of κ form again a lattice which we denote by $\partial\kappa$.

On the lattice, connections are represented by functions

$$g : E_\kappa \to G\,, \quad e \mapsto g_e \tag{4.1}$$

that map edges of κ into elements of the gauge group G. We denote the configuration space of all connections on κ by $\mathcal{A}_\kappa$. The gauge group is assumed to be a compact Lie group.

The quantities of physical interest are path integrals

$$W(\Phi) = \int \left(\prod_{e\in\kappa} \mathrm{d}g_e\right) \exp\Big(\iota\mathcal{S}(g)\Big)\Phi(g)\,. \tag{4.2}$$

For each edge e, we integrate over g_e by using the Haar measure on G. The action $\mathcal{S}$ is a real and gauge-invariant functional of the connection g, and Φ stands for a weighting functional. Depending on the context, the amplitude $W(\Phi)$ can be the partition function ($\Phi = 1$), the mean value of an observable Φ or a transition amplitude. ι is either i or -1, depending on whether we define a Euclidean or Minkowskian version of the theory.

We keep the choice of ι unspecified, but require the action to have the form

$$\mathcal{S} = \sum_{f\in\kappa^\circ} \mathcal{S}_f\,. \tag{4.3}$$

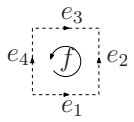


Figure 4.1: Example of edge orientations around a plaquette.

The sum runs over all faces in the interior κ° of the lattice. Each face action $\mathcal{S}_f$ is gauge-invariant and depends only on group elements of edges surrounding the face f.

The standard example is Euclidean SU(N) Yang-Mills theory in d dimensions, where the face action is chosen as the Wilson action

$$\mathcal{S}_f(g) = \frac{2N}{a^{4-d}\gamma^2}\left[1 - \frac{1}{2N}\,\mathrm{tr}\left(U_f(g) + U_f^\dagger(g)\right)\right]. \tag{4.4}$$

Here, U_f denotes the holonomy around the face. For face and edge orientations as in Fig. 4.1, it is given by

$$U_f(g) = g_{e_4}^{-1}\, g_{e_3}^{-1}\, g_{e_2}\, g_{e_1}\,. \tag{4.5}$$

γ is the gauge coupling and tr stands for the trace in the defining representation of SU(N). The classical continuum limit of (4.4) yields the Yang-Mills action

$$\mathcal{S} = \frac{1}{4\gamma^2}\int_M \mathrm{d}^d x\; F^a_{\mu\nu}F^{\mu\nu}_a\,. \tag{4.6}$$

In the quantum theory, one describes the limit $a \to 0$ by a sequence of lattice actions (4.4) that have an a-dependent gauge coupling $\gamma(a)$.

Another important example is lattice BF theory where the exponentiated face action equals

$$\exp\left(\mathrm{i}\,\mathcal{S}_f(g)\right) = \delta(U_f(g))\,. \tag{4.7}$$

Its definition can be formally derived from a continuum path integral

$$\int DA \int DB\; \exp\left(\mathrm{i}\,\mathrm{tr}\,[BF(A)]\right) \sim \int DA\,\delta(F(A))\,. \tag{4.8}$$

where, in addition to the connection degrees of freedom, we have a Lie algebra valued two-form field B.

In this article, we are primarily interested in amplitudes (4.2) of boundary states. By boundary states we mean weighting functionals Φ that depend only on group elements of boundary edges. We write the associated amplitude as

$$\Omega(\Phi) := \int\left(\prod_{e\in\kappa}\mathrm{d}g_e\right)\,\exp\left(\iota\mathcal{S}(g)\right)\Phi^*(\partial g)\,. \tag{4.9}$$

∂g denotes the restriction of the connection g to the boundary. We assume that Φ is an element in the Hilbert space $\mathcal{L}^2_0(\mathcal{A}_{\partial\kappa})$ of square-integrable and gauge-invariant boundary functionals, i.e.

$$\int\left(\prod_{e\in\kappa}\mathrm{d}g_e\right)\;\Phi^*(g)\Phi(g)\;<\;\infty\,. \tag{4.10}$$

Intuitively, one may think of $\Omega(\Phi)$ as the transition amplitude from "nothing" to the state Φ. When Φ is a product $\Phi_1\Phi_2$ of functionals, where Φ_1 and Φ_2 have disjoint domains, we can think of (4.9) as the transition amplitude between Φ_1 and Φ_2.

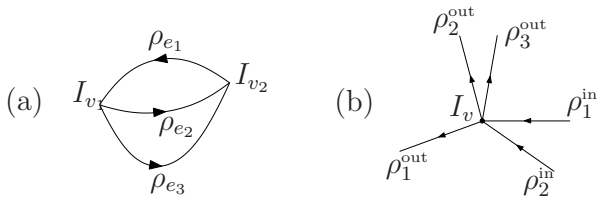


Figure 4.2: Labelling of spin network graphs.

4.3 Spin network states

A pure lattice gauge theory as described above can be transformed to a physically equivalent spin foam model. This so-called dual transformation rests on the use of a particular basis for functionals of connections: the basis of spin network states[2].

Generically, the term *spin network* refers to a directed graph of valence ≥ 2 that has a certain labelling (Fig. 4.2a): each edge e is labelled by a unitary irreducible representation ρ_e of the gauge group[3], and each vertex v carries an invariant tensor (or intertwiner) I_v in the tensor product

$$V_{\rho_1^{\text{out}}} \otimes \cdots \otimes V_{\rho_m^{\text{out}}} \otimes V^*_{\rho_1^{\text{in}}} \otimes \cdots \otimes V^*_{\rho_n^{\text{in}}} \,. \tag{4.11}$$

Here, each V_ρ stands for the representation space of the irrep ρ. For every outgoing edge e_i^{out} with label ρ_i^{out}, the tensor space has a component $V_{\rho_i^{\text{out}}}$, while for each incoming edge with label ρ_j^{in}, we get the dual of the associated representation space (Fig. 4.2b). We write $\rho = \text{triv}$ for the trivial representation. The representation labels are sometimes referred to as *colors* of edges.

To every spin network S corresponds a *spin network functional* or *state*, which we denote by Ψ_S. For a given connection g, the value $\Psi_S(g)$ of the functional is obtained as follows: for every edge e of the spin network graph, there is a group element g_e and its representation $\rho_e(g_e)$ in terms of the representation ρ_e. We contract all such representation maps with intertwiners, in the way indicated by the graph, and thereby receive a number—the value of S on the connection g. For later convenience, we enhance this rule by adding a factor $(\dim V_\rho)^{1/2}$ for every edge of the spin network. In formulas, we can write the value $\Psi_S(g)$ as

$$\Psi_S(g) = \left(\prod_{v \in \kappa} I_v \right) \cdot \left(\prod_{e \in \kappa} (\dim V_{\rho_e})^{1/2} \, \rho_e(g_e) \right) \tag{4.12}$$

where the dot $\cdot$ symbolizes the contraction of tensor indices.

In the present case, the connection lives only on the lattice κ, so spin network graphs lie on κ (Fig. 4.3). When the orientation of the spin network edge is opposite to that of the lattice edge, the spin network functional receives a factor $\rho_e(g_e^{-1})$ instead of $\rho_e(g_e)$.

So far, the correspondence between spin networks and spin network states is not one-to-one, since many different spin networks yield the same functional. It follows from the definition of the dual representation ρ^* on $V_{\rho^*} \equiv V^*_\rho$, for instance, that

$$I_2{}^{\cdots}{}_{\cdots a} \; \rho_e(g_e^{-1})^a{}_b \; I_1{}^{b \cdots}{}_{\cdots} = I_1{}^{b \cdots}{}_{\cdots} \; \rho^*_e(g_e)_b{}^a \; I_2{}^{\cdots}{}_{\cdots a} \,. \tag{4.13}$$

This means that we can reverse a spin network edge e, change the label from ρ_e to ρ_e^*, and still obtain the same spin network state (Fig. 4.4a). Likewise, a spin network with trivially

[2]The concept was first introduced in a context where the group was SU(2), hence the word "spin".

[3]More precisely, the ρ_e's are taken from a set $\mathcal{R}$ that contains one representative for each equivalence class of unitary irreducible representations of G.

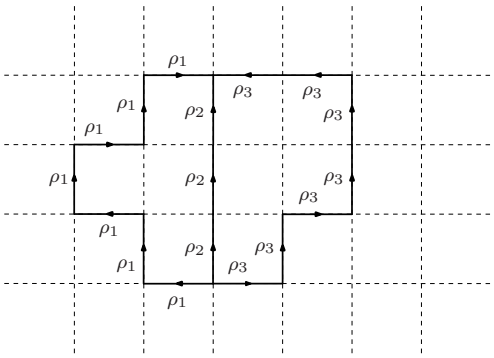


Figure 4.3: Spin network on a lattice (intertwiner labels are omitted).

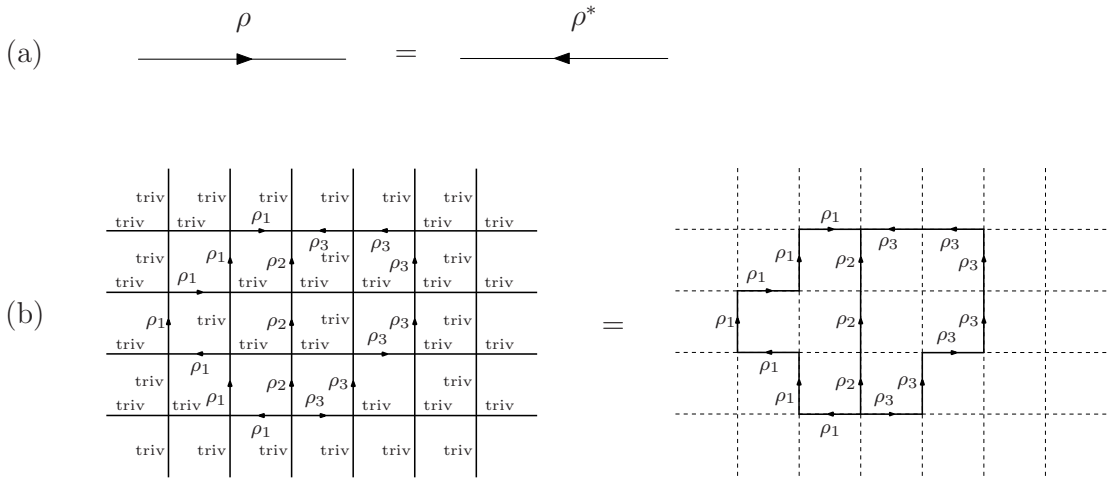


Figure 4.4: Equivalence of spin networks.

labelled edges defines the same functional as a spin network where these edges have been removed: representation tensors of the trivial irrep contribute just factors of 1 (Fig. 4.4b).

We take account of this redundancy and consider spin networks as *equivalent* when they define identical spin network functionals.

The use of the spin network states lies in the fact that they span the space $\mathcal{L}^2_0(\mathcal{A}_\kappa)$ of gauge-invariant functionals of the lattice connection. To proof this, one has to apply the Peter-Weyl theorem to each edge of κ [88]. Moreover, if we select orthonormal bases of intertwiners for the spaces (4.11), and take only these basis tensors as labels for spin networks, the resulting states form an orthonormal basis of $\mathcal{L}^2_0(\mathcal{A}_\kappa)$. We call this basis B_κ.

Analogously, we construct orthonormal bases for functionals over subgraphs of κ: for example, when we consider only spin networks on the boundary $\partial\kappa$, we obtain an orthonormal basis $B_{\partial\kappa}$ of $\mathcal{L}^2_0(\mathcal{A}_{\partial\kappa})$. The smallest admissible graphs consist of edges that surround a face f. The spin networks on such a graph are loops and provide an orthonormal basis B_f for functionals like the face action (4.4), which depend only on the connection around f.

Consider a loop spin network in B_f whose edges are coherently oriented as in Fig. 4.5. Then, all intertwiners are of the form $I_v \in V_{\rho_1} \otimes V^*_{\rho_2}$. Due to Schur's lemma, this intertwiner is only non-zero if $\rho_1 = \rho_2 = \rho$, and in that case, it is a multiple of the identity. Thus, we have only a single basis intertwiner in $V_\rho \otimes V^*_\rho$, and since it is normalized, it is fixed up to

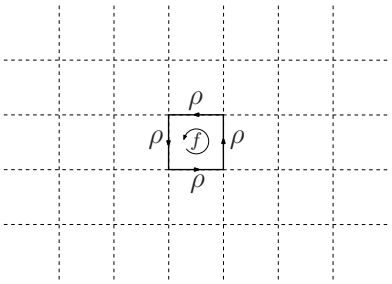


Figure 4.5: A loop spin network.

the choice of a phase: we take the phase factor to be one, so that

$$I^a{}_b = \frac{1}{(\dim V_\rho)^{1/2}}\,\delta^a{}_b\,. \tag{4.14}$$

When we insert this into the spin network functional, factors of dimension from edges and intertwiners cancel each other and we get

$$\begin{aligned}\Psi_{S_f}(g) &= \mathrm{tr}\left[\rho(g_{e_4}^{-1})\rho(g_{e_3}^{-1})\rho(g_{e_2})\rho(g_{e_1})\right] &(4.15)\\ &= \mathrm{tr}\left[\rho(U_f(g))\right]\,. &(4.16)\end{aligned}$$

That is, the loop functional is just the trace of the face holonomy in the irreducible representation ρ.

4.4 Dual transformation to the spin foam model

Let us now come to the actual dual transformation: for a given ultralocal lattice gauge theory, it maps the functional integral (4.2) over connections into a discrete sum over so-called *spin foams*: two-dimensional surfaces with branchings that have a certain labelling by irreps and intertwiners. Each spin foam is weighted by a factor—the spin foam amplitude—which is the analogue of the action and measure factors in the original theory. We can apply this transformation for every observable Φ, and thus obtain a new, but physically equivalent formulation of the gauge theory: connections are replaced by spin foams, action and measure by the spin foam amplitude, and observables Φ are translated into spin foam dependent weighting factors $\tilde{\Phi}$. We call this new theory a *spin foam model*.

The dual transformation of the path integral is achieved by expanding all functionals into spin network states, and by a subsequent integration over the connection. Here, we restrict ourselves to the case, where the amplitude is of the form (4.9). The procedure for more general amplitudes (4.2) is analogous.

It is convenient to split the evalutation of

$$\Omega(\Phi) = \int \left(\prod_{e\in\kappa} \mathrm{d}g_e\right)\, \exp\left(\imath\mathcal{S}(g)\right)\Phi^*(\partial g)\,. \tag{4.17}$$

into two steps. At first, we integrate the exponentiated action over group elements on the interior κ°: this leaves the connection on the boundary as a free variable, so we obtain a functional

$$\Omega(g) := \int\limits_{\partial g' = g} \left(\prod_{e\in\kappa^\circ} \mathrm{d}g'_e\right)\, \exp\left(\imath\mathcal{S}(g')\right)\,. \tag{4.18}$$

In the second step, we convolute this functional with the boundary state Φ, which yields the complete amplitude

$$\Omega(\Phi) = \int \left(\prod_{e \in \partial\kappa} \mathrm{d}g_e \right) \; \Omega(g) \Phi^*(g) \,. \tag{4.19}$$

Recall that the action decomposes into face actions, so (4.18) can be rewritten as

$$\Omega(g) = \int\limits_{\partial g' = g} \left(\prod_{e \in \kappa^\circ} \mathrm{d}g'_e \right) \prod_{f \in \kappa^\circ} \exp\left(\iota \mathcal{S}_f(g')\right) \,. \tag{4.20}$$

The expansion of $\exp\left(\iota \mathcal{S}_f(g)\right)$ into the spin network basis B_f gives

$$\exp(\iota \mathcal{S}_f) = \sum_{S_f \in B_f} c_{S_f} \Psi_{S_f} \,. \tag{4.21}$$

The subscript $S_f \in B_f$ means that we sum over all spin networks S_f whose states Ψ_{S_f} lie in the basis B_f. We see from this and equation (4.16) that both action and spin network states depend only on the face holonomy U_f, and that (4.21) is just another way of writing the character expansion

$$\exp\left(\iota \mathcal{S}_f(U_f)\right) = \sum_\rho \dim V_\rho \, c_{f\rho} \, \chi_\rho(U_f) \,. \tag{4.22}$$

In the following, we assume that for the trivial representation the coefficients $c_{f\rho}$ are 1. Otherwise we redefine the coefficients suitably.

By plugging (4.21) into the functional (4.20), we get

$$\begin{aligned} \Omega(g) &= \int\limits_{\partial g' = g} \left(\prod_{e \in \kappa^\circ} \mathrm{d}g'_e \right) \prod_{f \in \kappa^\circ} \sum_{S_f \in B_f} c_{S_f} \Psi_{S_f}(g) & (4.23) \\ &= \sum_{\{f\} \to \{S_f\}} \int\limits_{\partial g' = g} \left(\prod_{e \in \kappa^\circ} \mathrm{d}g'_e \right) \prod_{f \in \kappa^\circ} c_{S_f} \Psi_{S_f}(g) \,. & (4.24) \end{aligned}$$

In the second line, we pulled the summation symbol to the front, so we have to sum over all possible ways to assign basis spin networks S_f to the faces f of κ°.

Consider a single term in this sum, i.e. a summand for a fixed choice of basis spin network S_f for each face:

$$\int\limits_{\partial g' = g} \left(\prod_{e \in \kappa^\circ} \mathrm{d}g'_e \right) \prod_{f \in \kappa^\circ} c_{S_f} \Psi_{S_f}(g') \tag{4.25}$$

To determine the value of this term, we organize the appearing non-trivial loop spin networks into surfaces: two non-trivial loops belong to the same surface if they have only one common link and do not share this link with a third non-trivial loop. Different surfaces are either disconnected or they meet along links with other surfaces. The intersection of surfaces defines a graph which we call the *branching graph*: each line of this graph[4] must join at least three surfaces, otherwise their loops would form a single surface. Loops outside of surfaces carry the trivial representation. The ensemble of surfaces can be viewed as a *branched surface* F that consists of *unbranched components* F_i (see Fig. 4.6). We denote the branching graph by Γ_F.

[4]We refer to lines of the branching graph also as links, but one should not confuse them with links or edges of the lattice. A single link of the branching graph can be built from many lattice links.

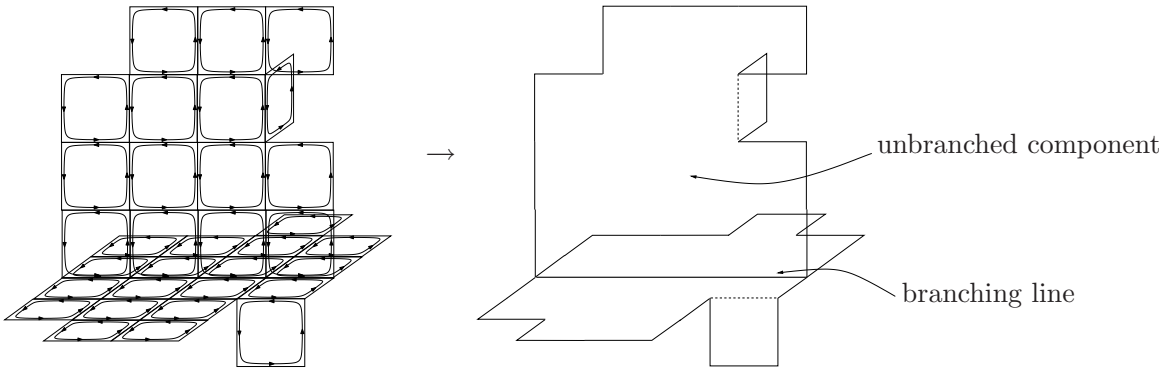


Figure 4.6: Grouping of loops into unbranched components.

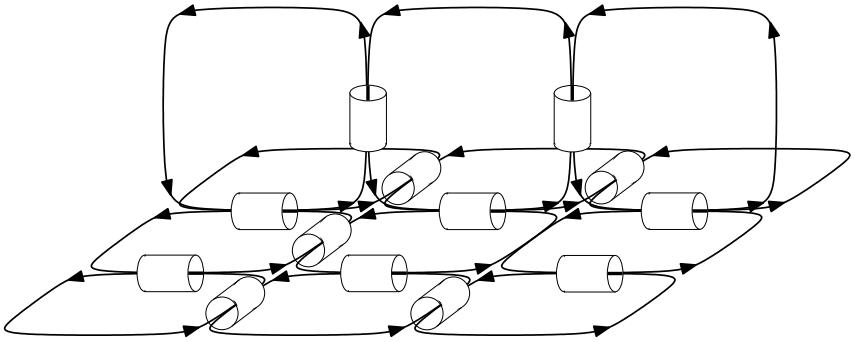


Figure 4.7: Representation of group integrations by cables (irrep and intertwiner labels are omitted).

Without loss of generality, we can choose representatives of loops that have coherent orientations within each unbranched component: i.e. orientations as in diagrams on Stoke's theorem. To represent the group integration, we use the diagrammatics of [89]: loop edges with label ρ_e symbolize the representation maps $\rho_e(g_e)$ and integrations over group elements are indicated by "cables" that surround the edges (see Fig. 4.7).

The next step consists in integrating out the group variables on each lattice edge e. When doing so, we have to distinguish between four different cases:

1. All incident loops on e are trivially labelled. In that case, the representation maps $\rho_e(g_e)$ are factors of 1 and the normalized Haar measure yields 1 as well. Therefore, all parts of the diagram that lie outside the branched surface contribute just 1 and disappear from the calculation.

2. Only one non-trivial loop is incident on e. Then, it follows from the identity

$$\int_G \mathrm{d}g \;\; \rho_1(g)^{a_1}{}_{b_1}\, \rho_2(g^{-1})^{b_2}{}_{a_2} = \frac{1}{\dim V_{\rho_1}}\, \delta^{a_1}{}_{a_2}\, \delta_{\ b_1}{}^{b_2}\, \delta_{\rho_1\,\rho_2}\,. \tag{4.26}$$

   that the integral is zero, since $\rho \neq_{\mathrm{triv}}$. This implies that an unbranched component is not allowed to end in a trivial part of the diagram, otherwise it gives zero in (4.25).

3. When exactly two non-trivial loops are incident on e, the representation labels have to be the same, due to (4.26). Thus, an unbranched component can only consist of loops with a single representation. We say that it is single-colored.

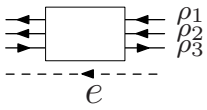


Figure 4.8: Example of a cable.

4. When e is shared by more than two non-trivial loops, the integral over G yields a so-called Haar intertwiner H. For a cable as in Fig. 4.8, we obtain, for example,

$$H^{a_1 a_2}{}_{a_3\, b_1 b_2}{}^{b_3} = \int_G \mathrm{d}g\; \rho_1(g)^{a_1}{}_{b_1}\, \rho_2(g)^{a_2}{}_{b_3}\, \rho_3(g^{-1})^{b_3}{}_{a_3} \tag{4.27}$$

$$= \int_G \mathrm{d}g\; \rho_1(g)^{a_1}{}_{b_1}\, \rho_2(g)^{a_2}{}_{b_3}\, \rho_3^*(g)_{a_3}{}^{b_3}\,. \tag{4.28}$$

   It is easy to see that tensors of this type are projectors onto invariant subspaces. In this example, H is a projector onto the invariant part of $V_{\rho_1} \otimes V_{\rho_2} \otimes V_{\rho_3}^*$.

Our reasoning shows that in (4.24) we only need to sum over certain configurations of loops—configurations that can be regarded as branched colored surfaces: each unbranched component is built from faces in the interior of the lattice and carries a non-trivial representation label or color. It also comes with an orientation that is determined by the coherent orientation of loops we started from. When an unbranched component has a boundary, the boundary links have to lie on the lattice boundary or on the branching graph where more than two components meet. An "open ending" is not allowed due to point 3.

To evaluate the contribution from a branched surface, we need to collect all factors and tensors that remain after the group integrations. We do this with the help of a graphical calculus that is close to that of [89]. Tensors and their contractions are represented by oriented labelled graphs: every vertex in the graph stands for a tensor T_v (indicated by a label T_v), while the edges between vertices symbolize the contraction of tensor indices. As in spin networks, the edges carry irreducible representations as labels. The irrep specifies the representation space for which the contraction is performed. At a vertex, the irreps of incoming and outgoing edges determine in which tensor space the tensor lives (cf. (4.11) and Fig. 4.2b). When an edge forms a closed curve, we interpret it as the trace in the respective representation. Note that two edges at a delta tensor are equivalent to a single edge without vertex:

(4.29)

With these conventions[5], the identity (4.26) takes the form

(4.30)

By "circle to the minus one" we mean the inverse of the dimension.

The "splitting identity" (4.30) is just a special case of that for the Haar intertwiner

$$H^{a_1 \cdots a_n}{}_{b_1 \cdots b_n} = \int_G \mathrm{d}g\; \rho_1(g)^{a_1}{}_{b_1} \cdots \rho_n(g)^{a_n}{}_{b_n}\,. \tag{4.31}$$

[5]Observe the difference to the rule for spin network functionals: in the case of spin networks, both vertices and edges are translated into tensors, while here the edges stand exlusively for contractions.

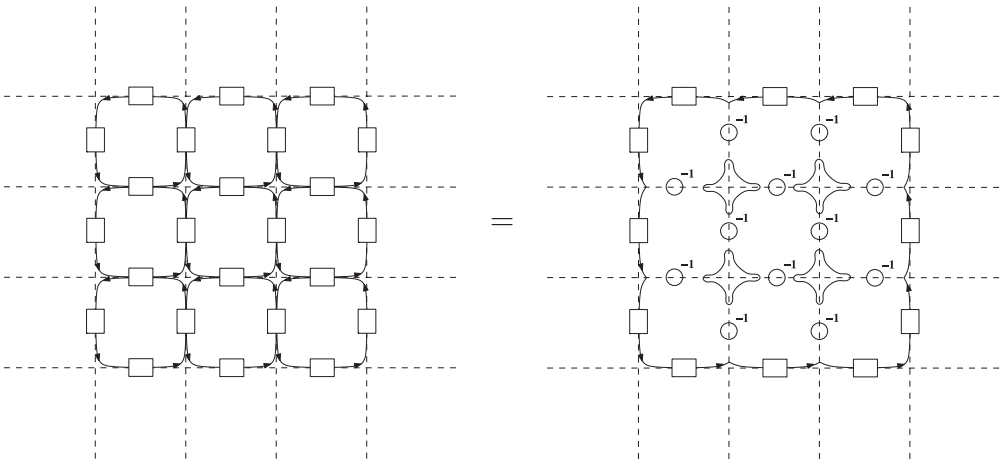
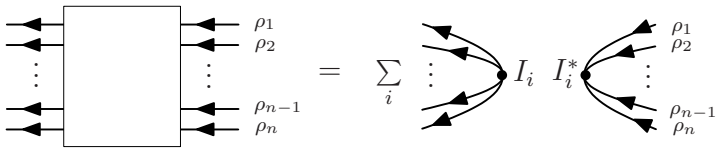


Figure 4.9: Integrations on a single-colored component.

in which case



(4.32)

The two diagrams are equal because H is a projector and can be written as a sum over basis intertwiners I_i:

$$H^{a_1\cdots a_n}{}_{b_1\cdots b_n} = \sum_i I_i{}^{a_1\cdots a_n}\, I_{i\, b_1\cdots b_n}\,. \tag{4.33}$$

Let us start by considering a single-colored component F_i. We see from diagram 4.9 that we get the dimension $\dim V_\rho$ for each vertex in the interior of F_i, and a factor $(\dim V_\rho)^{-1}$ for each interior edge. Each face contributes the loop coefficient

$$c_{S_f} = \dim V_\rho\, c_{f\rho} \tag{4.34}$$

(cf. eqns. (4.21) and (4.22)). When the unbranched component is without boundary, the factors accumulate to the amplitude

$$A_{F_i} = (\dim V_\rho)^{\chi(F_i)} \prod_{f\in F_i} c_{f\rho}\,. \tag{4.35}$$

Here, $\chi(F_i)$ stands for the Euler number of the unbranched component.

This formula is still true when F_i has a boundary. Since F_i is 2-dimensional, its boundary can only consist of components that have the topology of a circle. As shown in Fig. 4.9, the vertices and edges on the boundary fail to produce any dimensional factors. The missing factors, however, would just add up to

$$(\dim V_\rho)^{\chi(S^1)} = 1\,. \tag{4.36}$$

Therefore, equation (4.35) remains valid.

Next we analyze the contribution from the branching graph Γ_F: as explained in point 4., each lattice edge on a branching line produces a Haar intertwiner H. Between two vertices of Γ_F, all Haar intertwiners are the same, because the irrep labelling of incident loops does not change. Thus, we can use the projection property of H to replace the sequence of H factors by a single one (see Fig. 4.10).

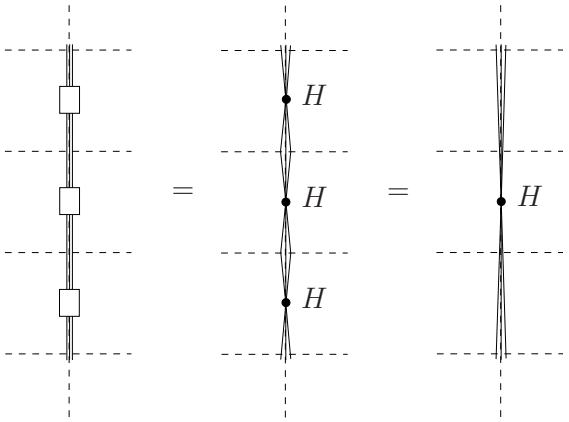


Figure 4.10: Haar intertwiners on branching line.

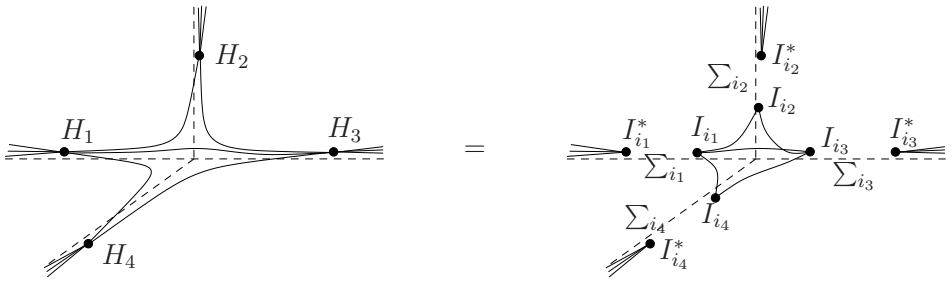
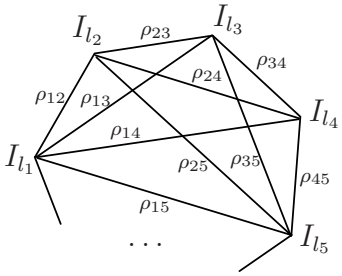


Figure 4.11: Splitting of Haar intertwiners.

After having done this for every branching line, we apply identity (4.32): each H is split into a sum over products of two basis intertwiners, which we "pull off" towards the vertices of the branching graph (Fig. 4.11). The result is a sum over possible assignments of basis intertwiners to links of Γ_F, and an amplitude

$$A_v = \quad (4.37)$$



for each vertex of the branching graph. The intertwiners come from links l_i of Γ_F that are incident on v, and the irreps ρ_{ij} belong to single-colored components that are bounded by pairs of links l_i, l_j.

Additional contributions arise when branching lines and unbranched components connect to the lattice boundary $\partial\kappa$: as depicted in Fig. 4.12, the splitting of Haar intertwiners "pushes" basis intertwiners out to the lattice boundary and the representation tensors $\rho_e(g_e)$ of boundary edges are left unintegrated. Therefore, what we get on the boundary is exactly a spin network functional! The associated spin network is determined by the branched surface F and can be visualized as a cut of the boundary through F (see Fig. 4.14): its intertwiners on vertices are fixed by the intertwiners on branching lines of F, and the colors on edges

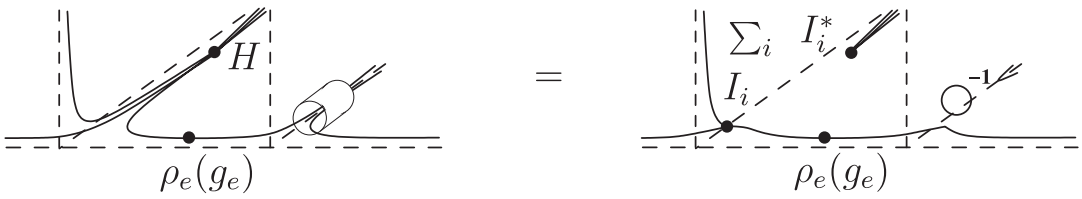


Figure 4.12: Branching line and single-colored component at the lattice boundary. The unintegrated representation tensor $\rho_e(g_e)$ is symbolized by a vertex.

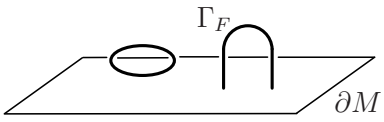


Figure 4.13: Degenerate examples of branching graphs.

equal the colors on unbranched components. We denote the boundary spin network by S_F and its functional by Ψ_{S_F}.

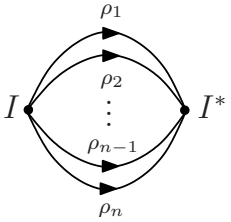
For later convenience, we want Ψ_{S_F} to be normalized: this is not yet the case, because the 2-valent intertwiners are just δ's and not $\delta/(\dim V_\rho)^{1/2}$, and the edges do not carry any factors $(\dim V_\rho)^{1/2}$. We add the needed factors to Ψ_{S_F}, and compensate this by absorbing their inverse into the amplitude of the single-colored surfaces F_i. The precise form of the correction factor depends on the topology of F_i : let B_{ij} denote the components of the boundary ∂F_i that lie on the lattice boundary $\partial\kappa$. For each component B_{ij} that is an open line, we receive a correction factor $(\dim V_\rho)^{-1/2}$, while we get just 1 when it is a loop. Thus, we can express the corrected amplitude for F_i as

$$A_{F_i} = (\dim V_\rho)^{\tilde{\chi}(F_i)} \prod_{f\in F_i} c_{f\rho} \tag{4.38}$$

where

$$\tilde{\chi}(F_i) = \chi(F_i) - \frac{1}{2}\sum_j \chi(B_{ij})\,. \tag{4.39}$$

For sake of completeness, we should mention that there are degenerate examples of branching graphs which we disregarded so far: a line of the branching graph can be closed or go directly from boundary to boundary (see Fig. 4.13). In that case, there is no vertex on the branching line, but nevertheless the splitting of Haar intertwiners leads to a diagram. It has the particularly simple form



. (4.40)

and just gives 1, since the basis intertwiners are normalized.

Let us summarize what we found: the functional $\Omega(g)$ can be written as a sum over terms which are characterized by branched surfaces F with a labelling: each unbranched component F_i of F carries an orientation and an irrep label, and every branching line comes with an intertwiner.

We call such a surface F, together with its labelling, a *spin foam* on the lattice κ. Remember that the initial choice of orientation on unbranched components was arbitrary. With the opposite sense, the representation label would be the dual, but everything else, including amplitudes, would remain unchanged. For that reason, we identify spin foams that are related by a reorientation and dualization of a single-colored component.

Given this definition, the functional $\Omega(g)$ is equal to a sum over all spin foams F on κ, where each spin foam contributes a spin network functional Ψ_{S_F} and an amplitude factor:

$$\Omega(g) = \sum_{F \subset \kappa} \left(\prod_{v \in \Gamma_F} A_v \right) \left(\prod_i A_{F_i} \right) \Psi_{S_F}(g) \,. \tag{4.41}$$

For each vertex of Γ_F, we get a factor of the form (4.37) and each single-colored component yields the amplitude (4.38).

Note that the amplitude of the branching graph is geometric in the sense that it only depends on the connectivity and labelling of the graph, but not on how its lines are discretized by edges of the lattice. The amplitude for single-colored components F_i contains also a geometric part, given by

$$(\dim V_\rho)^{\tilde{\chi}(F_i)} \tag{4.42}$$

which makes no reference to the lattice. In general, the product

$$\prod_{f \in F_i} c_{f\rho} \tag{4.43}$$

of face factors *does* depend on the discretization and constitutes the non-geometric part of the spin foam amplitude. It is only for special choices of $c_{f\rho}$ that the product (4.43) has a geometric interpretation. An example for this is BF-theory (see below) and the model we present in the next section.

In the final step, we contract the functional $\Omega(g)$ with the boundary state $\Phi(g)$ to give the value of the complete path integral (4.17). For that purpose, we expand $\Phi(g)$ in orthonormal basis spin networks:

$$\Phi(g) = \sum_{S \in B(\partial\kappa)} \Phi_S \, \Psi_S(g) \,. \tag{4.44}$$

When contracting (4.44) with (4.41), only those terms survive for which $S = S_F$, since Ψ_{S_F} is a basis spin network. Therefore, the final result is

$$\Omega(\Phi) = \sum_{F \subset \kappa} \left(\prod_{v \in \Gamma_F} A_v \right) \left(\prod_i A_{F_i} \right) \Phi^*_{S_F} \,. \tag{4.45}$$

For each spin foam F, the amplitude is multiplied by the coefficent of the boundary state with respect to the spin network induced by F on the boundary. Thus, we have attained a form for $\Omega(\Phi)$ that stands in analogy to the original path integral: the integration over connections has become a sum over spin foams, action and measure factors are replaced by the spin foam amplitude, and the weighting by the boundary functional $\Phi(\partial g)$ turns into the weighting by Φ_{S_F}.

By the same procedure, we can also transform amplitudes (4.2) for other observables Φ, and thereby translate all quantities of the lattice gauge theory into an equivalent theory that is purely formulated in terms of spin foams and spin networks. We refer to it as the spin foam model *dual* to the gauge theory.

The most simple spin foam model is that of BF theory (see eqns. (4.7) and (4.8)). The plaquette action has the character expansion

$$\delta(U_f) = \sum_\rho \dim V_\rho \, \chi_\rho(U_f) \tag{4.46}$$

so the plaquette coefficents $c_{f\rho}$ are all 1. The spin foam sum takes the form

$$\Omega(\Phi) = \sum_{F \subset \kappa} \left(\prod_{v \in \Gamma_F} A_v \right) \left(\prod_i (\dim V_{\rho_i})^{\tilde{\chi}(F_i)} \right) \Phi^*_{S_F} \tag{4.47}$$

Geometric spin foams

At this point, we should remark that the definition of spin foams that we employ here is not the same as the standard one appearing in the literature: usually, a spin foam is viewed as being built from *all* faces of the lattice, regardless of which labels they carry, and an amplitude is associated to every single vertex, edge and face. Here, we were led to merge sets of faces into more geometric objects that can be viewed as lying on the lattice (instead of *being* it), and the amplitudes of these objects depend only in part on the way they are subdivided by the lattice grid.

This organization of faces in "larger surfaces" was already emphasized in one of the first works on spin foams by Reisenberger [81], and it was known before in lattice gauge theory (see e.g. [1]). The same idea is also contained in Baez' definition of a category of spin foams [82].

To distinguish between the two notions of spin foam we call the ones defined here *geometric spin foams*: we can think of them as branched labelled surfaces in the continuous manifold M that are independent of any choice of lattice, similarly as a continuous field does not rely on a cutoff. From that point of view, the lattice regularization becomes the requirement that one should not sum over all possible geometric spin foams, but only over those that can be "fitted" onto the lattice κ: by that we mean that every single-colored component is a union of plaquettes of κ, and every branching line is a union of links of κ.

So far the lattice is just a means to regularize the theory, but it has no deeper conceptual justification. In the spin foam approach to gravity, one hopes to explain the spacetime lattice itself as being a consequence of spin foams. If that was true, we would arrive at a picture where spin foams of the gauge theory are placed on spin foams of gravity, and no external background structure or cutoff has to be imposed anymore. We will say more about this in section 4.6.

>From a space *plus* time point of view, spin foams are interpreted as evolving spin networks [37]: when the boundary spin network consists of two disconnected spin network graphs, the interpolating spin foam can be viewed as the worldsheet of the transition from one spin network to the other (see Fig. 4.14). In the process, vertices can blow up to subgraphs, and subgraphs may shrink to a vertex. When there is only a single connected spin network, we regard the spin foam as a creation or annihiliation process.

4.5 A spin foam model of Yang-Mills theory

In this section, we describe a spin foam model that is dual to Euclidean lattice Yang-Mills theory. We take SU(N) as the gauge group. The spacetime dimension d can have any value ≥ 2.

As regards the choice of the face action, we have several possibilities, all of which produce the same classical continuum limit. The Wilson action (4.4) appears as the most simple

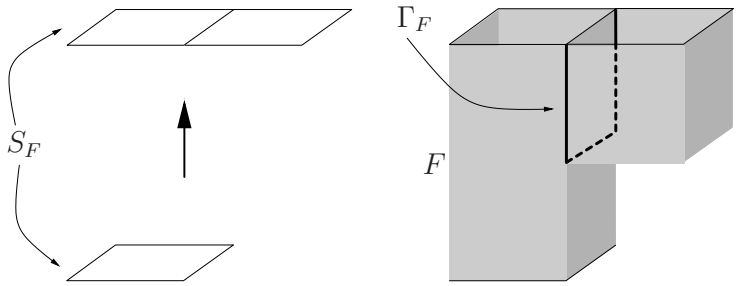


Figure 4.14: Spin foams as worldsheets of spin networks: the branching graph Γ_F (thick line) indicates where surfaces with different irrep labels meet.

choice on the level of group variables, but under the dual transformation it leads to relatively complicated face coefficients $c_{f\rho}$. In that respect, the most simple alternative is the heat kernel action, which has the character coefficients

$$c_{f\rho} = \exp\left(-a^{(d-4)}\gamma^2(a)\, C_\rho\right) . \tag{4.48}$$

Here, C_ρ is the eigenvalue of the Casimir operator in the ρ-representation. On the side of the lattice field theory, this action takes the form

$$\exp\left(-S_f(U_f)\right) = \frac{K\left(U_f, \frac{\gamma^2}{2}\right)}{K\left(\mathbb{1}, \frac{\gamma^2}{2}\right)} . \tag{4.49}$$

The heat kernel K is determined by the differential equation

$$\frac{\partial}{\partial t}K(U,t) = \Delta K(U,t)\,, \qquad K(U,0) = \delta(U)\,. \tag{4.50}$$

Δ denotes the Laplace-Beltrami operator on the group.

In the following, we take (4.49) as the starting point for the dual transformation. The Wilson and heat kernel action are believed to be in the same universality class [90], so the preference of one or the other should not affect the low-energy physics of the theory.

For a single-colored component F_i with label ρ_i, the factors (4.48) add up to

$$\prod_{f\in F_i} c_{p\rho_i} = \exp\left(-N_{F_i}\, a^{(d-4)}\gamma^2(a)\, C_{\rho_i}\right) , \tag{4.51}$$

N_{F_i} stands for the number of plaquettes in the surface F_i. Clearly, N_{F_i} is proportional to the area $\mathcal{A}_{F_i}$ of the single-colored component, so we can write

$$N_{F_i} = \frac{\mathcal{A}_{F_i}}{a^2} \tag{4.52}$$

and

$$\prod_{f\in F_i} c_{p\rho_i} = \exp\left(-\mathcal{A}_{F_i}\, a^{(d-6)}\gamma^2(a)\, C_{\rho_i}\right) . \tag{4.53}$$

The area is accompanied by factors which we absorb into the new constant

$$T_\rho(a) := a^{(d-6)}\gamma^2(a)\, C_{\rho_i}\,. \tag{4.54}$$

The overall amplitude for a boundary state Φ becomes

$$\Omega(\Phi) = \sum_{F\subset\kappa} \left(\prod_{v\in\Gamma_F} A_v \right) \left(\prod_i (\dim V_{\rho_i})^{\tilde{\chi}(F_i)} \exp\Big(- T_{\rho_i} \mathcal{A}_{F_i} \Big) \right) \Phi^*_{S_F} \,. \tag{4.55}$$

where the vertex amplitudes are defined as in equation (4.37).

As a result of the particular choice for the face coefficients (eqn. (4.48)), we arrive at a model where the amplitude of a spin foam depends only on the geometric properties of the spin foam—its topology, labelling and area. The lattice regularization enters only through the value of $T_\rho(a)$ and the requirement that spin foams should be congruent with the lattice.

The form of the sum (4.55) bears a striking resemblance to lattice quantizations of the Nambu-Goto string [91]: the spin foams play the role of worldsheets for the spin networks, and the exponent can be viewed as an action for each single-colored sheet that is proportional to its area. Thus, we interpret the factor (4.54) as a color-dependent tension T_ρ of spin network edges. The running of the original gauge coupling $\gamma(a)$ is mapped into a running of the edge tension.

4.6 Background independent spin foam models

We have seen in the previous section that the notion of geometric spin foams allows for a remarkably simple dual formulation of Yang-Mills theory. In this section, we will argue that it is also particularly suited for constructing background-independent theories.

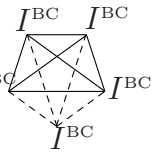
In the spin foam approach to gravity, one defines models that are no longer dual transforms of pure gauge theories. In many cases, however, their definition is closely related to the duality map, as they can be obtained by modifiying the spin foam model dual to BF theory. The most prominent example is the Barrett-Crane model for Riemannian or Lorentzian gravity [92]: for its heuristic derivation, we can start from a formal path integral over the tetrad and connection field. This path integral is rewritten as a BF path integral

$$\int DA \int DB \, \exp\Big(\mathrm{i}\,\mathrm{tr}\,[BF(A)] \Big) \tag{4.56}$$

with additional constraints on the B-field. One introduces a triangulation of spacetime, transforms pure BF theory into its dual spin foam model, and translates the constraints on B into restrictions on the spin foams [44].

In the case of the Riemannian Barrett-Crane model, the spin foam labels are restricted to balanced[6] representations of SO(4), and the intertwiners are replaced by so-called Barrett-Crane intertwiners I^{BC}. If we denote such spin foams by F^{BC}, we can write the entire spin foam sum as

$$\Omega(\Phi) = \sum_{F^{\mathrm{BC}}\subset\kappa} \left(\prod_{v\in\Gamma_{F^{\mathrm{BC}}}} \right. \quad \left. \right) \left(\prod_i (\dim V_{\rho_i})^{\tilde{\chi}(F^{\mathrm{BC}}{}_i)} \right) \Phi^*_{S_{F^{\mathrm{BC}}}} \,. \tag{4.57}$$



The lattice κ consists of the dual complex of the chosen triangulation. The vertices of the branching graph have valence 4 or 5. Accordingly, the number of intertwiners in a vertex amplitude varies between 4 and 5 (indicated by dashed lines in (4.57)). We see that the transition from BF-theory to (4.57) preserved the geometric form of the sum—the spin foam

[6] A balanced representation is isomorphic to a tensor product $j\otimes j^*$ where j is an irreducible represenation of SU(2).

amplitude has changed, but it is still a function of the geometry of the spin foam. There are other versions of the Barrett-Crane model that differ from (4.57) in that they contain additional modifications of the amplitude. In general, this breaks the topological invariance of surface amplitudes and makes our geometric interpretation impossible. We show in the appendix that (4.57) corresponds to version B in [93].

In either case—whether we regard spin foams as consisting of the entire lattice, or as geometric objects that are placed on it—the construction rests on the choice of a particular lattice κ, and that clashes with the idea of defining a theory that is independent of any background structure.

One way to resolve this problem goes along with the viewpoint that identifies spin foams with the lattice: the lattice itself is interpreted as a discrete spacetime, and the sum over spacetime fluctuations is then implemented as a sum over a large class of lattices (and their labellings). Thus, by summing over lattices, one avoids the choice of any particular κ. The precise form of this sum is obtained from the perturbative expansion of a group field theory[7] [83, 84].

Here, we take a different approach, based on the philosophy that spin foams are lattice-independent geometric entities: we view the lattice just as an auxiliary background which was used for deriving the amplitudes, and discard it when defining the full background independent model. For this to be possible, we need to start from models whose amplitudes depend only on the geometry of spin foams, and not on the lattice on which the spin foam is defined. In the absence of a background metric, such amplitudes can only depend on topological properties—the topology of single-colored sheets and how they are connected by the branching line.

Another way of stating this condition is to say that spin foam amplitudes must be topological invariants. Keep in mind that when we use the word "topological" in this way, it only refers to amplitudes of single spin foams, but it has nothing to do with topological invariance of the entire spin foam sum. The amplitudes of spin foams could be topological in the sense we just mentioned, and yet give a model that is not topologically invariant.

Consider a spin foam sum

$$\Omega_\kappa(\Phi) = \sum_{F\subset\kappa} \left(\prod_{v\in\Gamma_F} A_v\right)\left(\prod_i A_{F_i}\right)\Phi^*_{S_F} \tag{4.58}$$

that satisfies our requirement: each spin foam amplitude is uniquely determined by the branching graph, the coloring and topology of single-colored sheets. Assume also that the boundary functional Φ has a topological dependence: the coefficient Φ_{S_F} is a function of the connectivity and labelling of the spin network S_F, and no information on the lattice is needed to compute its value. In that case, we extend (4.58) to a background independent sum over all spin foams in the manifold:

$$\Omega(\Phi) = \sum_{F\subset M} \left(\prod_{v\in\Gamma_F} A_v\right)\left(\prod_i A_{F_i}\right)\Phi^*_{S_F} \tag{4.59}$$

(The same idea could be also expressed by a refinement limit: equip the manifold with an auxiliary Euclidean metric, and choose a sequence κ_i of triangulations for which the volume of tetrahedra goes uniformly to zero. The formal limit

$$\lim_{i\to\infty} \Omega_{\kappa_i}(\Phi) \tag{4.60}$$

contains all spin foams on M whose valence does not exceed that of a dual triangulation.)

[7] a generalization of matrix theories

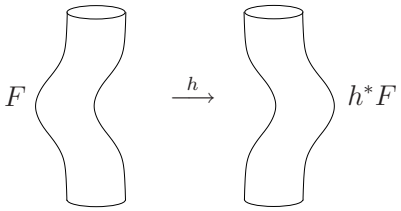


Figure 4.15: Action of a homeomorphism h on a spin foam F.

The transition from (4.58) to (4.59) requires a generalization of the Hilbert space, since the spin networks S_F are no longer restricted to the boundary lattice $\partial\kappa$. We define the new space of boundary states as follows: we take the space

$$\mathcal{H}_{\partial M} := \left\{ \sum_{i=1}^{n} a_i S_i \;\Big|\; a_i \in \mathbb{C}\,,\;\; S_i \subset M\,,\;\; n \in \mathbb{N} \right\} \tag{4.61}$$

of finite linear combinations of spin networks in the boundary ∂M, equip it with the inner product

$$\langle S, S' \rangle = \delta_{S,S'}\,. \tag{4.62}$$

and define the space of boundary states as $\mathcal{H}^*_{\partial M}$, i.e. as the space of linear functionals $\Phi : \mathcal{H}_{\partial M} \to \mathbb{C}$.

Clearly, the spin foam sum (4.59) contains a huge overcounting, which is due to the fact that amplitudes do not depend on how spin foams are embedded in the manifold (Fig. 4.15). More precisely, each spin foam amplitude

$$A(F) := \left(\prod_{v \in \Gamma_F} A_v \right) \left(\prod_i A_{F_i} \right) \tag{4.63}$$

and boundary coefficient Φ_{S_F} is invariant under the action of homeomorphisms $h : M \to M$, which we write as

$$A(h^*F) = A(F) \qquad \text{and} \qquad \Phi_{S_{h^*F}} = \Phi_{S_F}\,. \tag{4.64}$$

To eliminate this overcounting, we factor off infinite gauge volumes à la Fadeev-Popov, and replace (4.59) by a sum over equivalences classes $\tilde{\mathcal{F}}$ of spin foams under homeomorphisms:

$$\Omega(\Phi) \;=\; \sum_{\tilde{\mathcal{F}}} V(\mathrm{Homeo}(M))\, \tilde{A}(\tilde{\mathcal{F}})\, \tilde{\Phi}^*_{\tilde{S}_{\tilde{\mathcal{F}}}} \tag{4.65}$$

$$\downarrow$$

$$\tilde{\Omega}(\tilde{\Phi}) \;=\; \sum_{\tilde{\mathcal{F}}} \tilde{A}(\tilde{\mathcal{F}})\, \tilde{\Phi}^*_{\tilde{S}_{\tilde{\mathcal{F}}}}\,. \tag{4.66}$$

We call such equivalence classes *abstract* or *topological* spin foams. Correspondingly, we define an *abstract spin network* as an equivalence class of spin networks under homeomorphisms of the boundary. $\tilde{S}_{\tilde{\mathcal{F}}}$ stands for the equivalence class of spin networks that is induced by $\tilde{\mathcal{F}}$. The functionals $\tilde{A}$ and $\tilde{\Phi}$ are defined by

$$\left.\begin{array}{l} \tilde{A}(\tilde{\mathcal{F}}) = A(F) \\ \tilde{\Phi}(\tilde{S}_{\tilde{\mathcal{F}}}) = \Phi_{S_F} \end{array}\right\} \text{ for any representant } F \text{ of } \tilde{\mathcal{F}}. \tag{4.67}$$

In a more explicit form, equation (4.66) reads

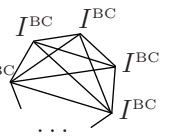
$$\tilde{\Omega}(\tilde{\Phi}) = \sum_{\tilde{\mathcal{F}}} \left(\prod_{v \in \tilde{\Gamma}_{\tilde{\mathcal{F}}}} A_v \right) \left(\prod_i A_{\tilde{\mathcal{F}}_i} \right) \Phi^*_{\tilde{S}_{\tilde{\mathcal{F}}}} \tag{4.68}$$

where each tilded quantity is an equivalence class under homeomorphisms.

The definition of abstract spin foams and spin networks parallels that of abstract spin networks in canonical loop quantum gravity[8], and that of 3- and 4-geometries in the sum-over-metrics approach to gravity. In our case, the dynamics is insensitive to moduli of spin networks, so we extended the symmetry group from diffeommorphisms to homeomorphisms. As defined above, abstract spin foams are closely related to the spin foams of group field theories, which are labelled non-embedded 2-complexes. There is a crucial difference, however: the latter can have trivial labels, while, by construction, our spin foams carry only non-trivial representations. We discuss the consequences of that difference in the final section.

The upshot of all this is the following: by going through the steps from (4.58) to (4.68), we can start from any spin foam model on a lattice whose amplitudes satisfy topological invariance, and construct a manifestly background free theory from it. The latter is specified by sums over abstract spin foams, which carry only topological and combinatorial information. There is, in particular, a version of the Barrett-Crane model which meets our requirements (see (4.57)), so we can extend it to an abstract sum

$$\tilde{\Omega}(\tilde{\Phi}) = \sum_{\tilde{\mathcal{F}}^{\mathrm{BC}}} \left(\prod_{v \in \tilde{\Gamma}_{\tilde{\mathcal{F}}^{\mathrm{BC}}}} \right. \quad \left. \right) \left(\prod_i (\dim V_{\rho_i})^{\tilde{\chi}(\tilde{\mathcal{F}}^{\mathrm{BC}}{}_i)} \right) \Phi^*_{\tilde{S}_{\tilde{\mathcal{F}}^{\mathrm{BC}}}} \,. \tag{4.69}$$



As such, a sum of the type (4.69) is highly divergent: for any given abstract spin network in the boundary, there appears an infinite number of disconnected spin foams, an infinity of topologies for each single-colored sheet, and also infinitely many connected spin foams, due to the presence of bubbles [94]. To arrive at a concrete model, an appropriate truncation or dampening will be necessary: for example, a restriction to connected spin foams, a cutoff on topologies and an exclusion of bubbles.

4.7 Summary and discussion

Let us summarize the contents and results of the paper:

We have given a pedagogic derivation of the transformation from pure lattice gauge theory to its dual spin foam model. In doing so, we emphasized the grouping of plaquettes into single-colored surfaces, and were naturally led to a geometric definition of spin foams: spin foams are not identified with the lattice and its labellings, but instead regarded as geometric objects—branched surfaces—that are placed on the lattice. This geometric viewpoint enabled us to write down a very simple spin foam model of lattice Yang-Mills theory for gauge group SU(N) and dimension $d \geq 2$. Its spin foams are weighted with an "action" that is proportional to the area of surfaces, similar as worldsheets of the Nambu-Goto string. The proportionality constant depends on the representation label of unbranched sheets and can be viewed as a tension of spin network edges. The running of the original gauge coupling is mapped into a running

$$T_\rho(a) := a^{(d-6)} \gamma^2(a)\, C_\rho$$

[8] See e.g. sec. 6.4 in [11]

of the edge tension. It should be stressed that the transformation from gauge theory to spin foam model is non-perturbative, so it does not require a strong coupling expansion.

In the second part of the article (sec. 4.6) we applied the notion of geometric spin foams to models of gravity: we introduce a symmetry condition that requires spin foam amplitudes to be independent of the lattice. There are two motivations for this step: firstly, some of the existing models meet the symmetry requirement, or do so after a simple modification of the amplitude. Secondly, it allows for a purely geometric definition of the spin foam sum: for each spin foam, the amplitude depends only on the topology and labelling of the branched surface, and the lattice cutoff translates into the condition that we should only sum over spin foams that lie on the lattice. In models that have the desired symmetry property, we discard the lattice regularization and extend the sum to all spin foams in the manifold. After factoring off infinite gauge volumes (the volume of the homeomorphism group), we arrive at a sum over abstract spin foams. Thus, we obtain a purely combinatorial model that is free of any choice of lattice or triangulation. Our procedure applies to a version of the Barrett-Crane model, as we show in the appendix.

The two parts of the article are tied together by the idea that spin foams should be regarded as geometric objects. Strictly speaking, however, the gauge and gravity case are treated on a different footing: in Yang-Mills theory, we picked out a dual model that has a nice geometric interpretation, but that does not prevent us from using other non-geometric models (e.g. the model associated to the Wilson action) to describe the same low-energy physics. For the gravity models, on the other hand, we impose the symmetry condition from the start, and exclude models that do not satisfy it. We do that, because a lattice dependence of amplitudes would go against the idea of background independence, so we need amplitudes that are only determined by geometric properties of the spin foam.

Yang-Mills spin foam model

The construction of the Yang-Mills model rests on two inputs: the simplicity of the character coefficients of the heat kernel action, and the combination of plaquettes into larger single-colored surfaces. Although both of this is known in lattice field theory and quantum gravity, we have not found any previous definition of the model in the literature. It would be interesting to include fermions and see how they add into the geometric picture of the model. Could it provide an alternative method in the non-perturbative analysis of QCD? A great deal of research has been devoted to the relation between Yang-Mills theories and string theory: there is the "old" idea of describing lattice gauge theory in terms of an effective string theory[9], and the more recent program on the correspondence between supersymmetric Yang-Mills theory and superstring theory [96]. As we have seen, our dual model shares features with the Nambu-Goto string, so can it tell us anything about the string-gauge theory relation? What is its large N limit?

Another interesting question is if the model could be combined with gravity spin foams to give a coupling of Yang-Mills theory and gravity. That possibility is suggested by the area weighting of the Yang-Mills spin foams and the interpretation of gravity spin foams as quanta of area.

Geometric spin foams versus group field theory

In the context of gravity, our geometric viewpoint suggests a natural way to overcome the triangulation dependence of spin foam models. We consider only models whose weights are functions of topological properties of the spin foam: i.e. we have a rule that determines the

[9]For references, see e.g. [95].

amplitude for a given branching graph, coloring and topology of surfaces, without making reference to the underlying triangulation. When a model has this property, we keep the rule for the amplitudes, but replace the sum over spin foams on the triangulation by a sum over abstract spin foams—topological equivalence classes of spin foams. The resulting new model is manifestly background-independent and contains no information on the triangulation we started from.

Thus, our approach provides an alternative to the group field theory method, where the triangulation dependence is removed by summing over a large class of complexes. The sum over complexes results from a sum over Feynman diagrams in the perturbative expansion of a group field theory (GFT). The graph of a Feynman diagram corresponds to an abstract 2-complex, and its value is a spin foam sum on the complex, i.e. a sum over irrep labels on faces and intertwiners on edges. In this context, a spin foam is identified with the labelled, abstract complex, so we can view the entire expansion as a sum over abstract spin foams.

What is the relation between this sum and the sum over abstract spin foams we have defined? The answer is that they are very different, and the reason for it lies in different definitions of spin foams: in the GFT sense, a spin foam is an abstract, non-embedded 2-complex, together with irrep labels on faces and intertwiner labels on edges. Trivial irreps are *allowed*.

In our approach, we also arrived at abstract spin foams, but we started from the geometric concept of spin foam, where faces of the same color are merged into larger surfaces, and trivial labels are ignored. Therefore, our abstract spin foams are equivalent to non-embedded 2-complexes where trivial irreps are *excluded*.

This difference has dramatic consequences for the spin foam sum: in the GFT expansion, one sums over all labelled complexes, no matter how much or little of the complex is labelled trivially. In particular, the same non-trivial part can appear in infinitely many spin foams which just result from adding trivial parts to it. The weights of these spin foams differ only by symmetry factors and powers of the coupling constant, while the actual spin foam amplitude stays the same. In our approch, the non-trivial part is only counted once, since trivial labels do not appear in the bookkeeping.

For that reason, the present proposal for abstract spin foam sums seems better suited for an interpretation in terms of histories of spin networks: here, a history of non-trivial spin networks is counted once, while the GFT counts also all ways of adding trivial spin networks to it.

Semiclassical analysis

A central problem in the research on spin foam models is the semiclassical limit. We would like to know if theories like the Barrett-Crane model can produce any physically realistic low-energy limit. Can it generate backgrounds that resemble gravity? Does it exhibit critical behaviour? Following the ideas in [70], we can sketch a possible way to tackle these questions: we view abstract spin foams as classical configurations, and the associated amplitudes as a kind of exponentiated action $\exp(\mathrm{i}\,\mathcal{S})$. In the space of abstract spin foams, we have a certain notion of continuity, which is given by incremental relabelling and stepwise modification of spin foam topologies. Thus, we can define variations of the action with respect to spin foam configurations. The analogy with field theory suggests that backgrounds might be identified as large spin foam complexes that are extrema of the action. If such extrema exist, one could try to formulate perturbative expansions around them.

Chapter 5

Summary of thesis

5.1 Evaluation of results

Let us give a compact summary of all the results of the thesis (for more details, we refer the reader to the summaries of chapters 2, 3 and 4):

The work of chapter 2 was motivated by the question if the general boundary idea could be concretized in a simple example, namely, that of free scalar field theory in a Euclidean flat background. We have managed to define a discretization of the path integral with general boundaries. Based on a number of assumptations, we could derive a generalized Schrödinger equation from it, which governs the dependence of the path integral on deformations of the boundaries.

This needs to be completed in several ways in order to be considered as a mathematical implementation of the general boundary formalism. In that sense, our result is not satisfying: what is missing is a prescription for Wick rotation to the Lorentzian theory, which is the physically interesting case, and a proof that the discretized path integral has a well-defined continuum limit.

In chapter 3, we have found a physically motivated way to translate the vacuum of linearized and regularized ADM gravity into a state of the LQG Hilbert space. The resulting state is a Gaussian superposition of spin networks and has its peak at a so-called weave—a type of state that already appeared earlier in the literature. The more a spin network differs from the weave in its graph and labelling, the more it is damped in the superposition. These properties can be traced back to the semiclassical peakedness of the ADM state, and therefore be interpreted as the way in which semiclassicality manifests itself at the level of spin networks. Accordingly, we interpret the peak weave as the spin network *pendant* of the classical background, and the spin networks near this peak as the semiclassical fluctuations around it. This notion of semiclassicality can be very helpful in conceiving methods for a semiclassical approximation, and at the end of the article we sketched ideas on how this might be used to derive a perturbative expansion of the scalar constraint of LQG.

In addition, we have found the interesting property that the peak of the state becomes independent of the regulating parameter when the latter goes to zero. This is an indication of the suspected self-regulating property of LQG: the degrees of freedom of the quantum theory (i.e. the spin networks) are such that fluctuations with momenta higher than the Planck length cannot be represented on them and drop out. As a result, the peak of the state remains the same, although the regulating momentum cutoff Λ is sent to infinity.

Let us mention that our vacuum and graviton states have been used in a recent article by Modesto and Rovelli, in attempts to compute graviton scattering for a spin foam model [97].

At last, we come to the article on spin foam models (chapter 4): we have recast the dual

transformation from lattice gauge theory to spin foam model in a form that emphasizes a geometric viewpoint on spin foams: i.e. spin foams are not identified with the lattice, but instead regarded as geometric objects that lie on it. This viewpoint allows us to identify an interesting property of the spin foam model of 3d gravity and the Barrett-Crane model of 4d gravity: the amplitudes of a single spin foam do not depend on the lattice on which it is defined. This means that the lattice regularization enters only through the requirement that spin foams should lie on the lattice. We propose that this property of lattice independence should be elevated to a symmetry requirement for any spin foam model that is supposed to give a background-independent theory.

For models which satisfy this condition, it is easy to make the transition to a lattice-independent model. We simply drop the requirement that the sum should be restricted to spin foams on the lattice, and after factoring off an infinite overcounting, we arrive at a sum over abstract spin foams—equivalence classes of spin foams under homeomorphisms. The reference to a lattice has disappeared.

We think, firstly, that this proposal brings the spin foam approach closer to its aim of defining truly background-independent models, and, secondly, that it renders it more accessible to a semiclassical analysis: now, one is no longer confused by the presence of an arbitrarily chosen lattice or triangulation, but able to concentrate on the actual physical configurations, which are the abstract spin foams.

As a side result, we obtained a simple dual formulation of lattice Yang-Mills theory that is reminiscent of the Nambu-Goto string on a lattice: the single-colored surfaces of the spin foams are weighted with the exponentiated area. This reformulation might be interesting for both string and lattice gauge theorists.

5.2 Critical assessment of LQG and spin foam approach

Before concluding, we make some critical remarks on the present state of the research on canonical LQG and spin foam gravity. We concentrate on the *negative* aspects, since they indicate what needs to be done to improve the theory.

Dynamics in canonical LQG

In the formulation of canonical LQG, the most urgent problem seems to be a definition of the dynamics, i.e. an implementation of the scalar constraint or finite evolutions as operators. There has been the important result of Thiemann, who managed to translate the constraint of classical Ashtekar-Barbero gravity into an operator on the dual Cyl^* of the cylindrical functions. This construction has several problems that we briefly discussed in the introduction (see section 1.4.6): we think that taking the ϵ-limit of the dual operator has no apparent physical motivation. Moreover, it seems unnatural to use a triangulation for regulating the operator that is finer than the spin network graph. If we view spin networks as the objects that *constitute* space, it would appear more appropriate to take the spin networks themselves as the "triangulation" on which one defines the scalar constraint. This might also resolve the problem of ultralocality, since then an ultralocal action of the scalar constraint operater on spin networks could be avoided. Note that this idea is related to the viewpoint in chapter 4, where we take the lattice-independent spin networks and spin foams as the central objects, and the initially present lattice drops out of the description.

It has been argued that anomaly-freeness of the constraint operator algebra could provide a crucial check on the dynamics of LQG [15]. However, since a definition of the scalar constraint (or finite evolutions) is missing, a computation of anomaly terms is not yet in reach.

Semiclassical limit

The big unresolved problem of both canonical LQG and spin foam gravity is the question of whether they correspond to physically realistic effective field theories in certain regimes.

As mentioned in the introduction, the often expressed viewpoint is that LQG is a quantization of Einstein gravity, without the introduction of new type of degrees of freedom. As a result, a lot of emphasis is laid on the careful quantization procedure that is supposed to ensure that we still deal with gravity after quantization. At the same time, very little is known about the actual physical properties of these systems.

We argue that one can also take an alternative viewpoint, namely, that the construction of LQG takes the classical Einstein gravity as a *guiding principle* for arriving at a more fundamental new theory that is not a field theory anymore, though related to it. From that point of view, quantization should not be taken as a fixed procedure, but rather as a means to *guess* what the degrees of freedom of the fundamental theory are, and how its dynamics should be defined. If we take that position, the decisive criterion will be whether our models have physically interesting properties, and it will be less important if they are the precise result of some quantization rule.

We think that such an attitude could help in finding new ways of formulating the theory, and, in particular, encourage simplifying changes in the definition of the dynamics. So far, both in canonical LQG and spin foam gravity, the models are the result of a quantization procedure and have a rather complicated definition of the dynamics. If we shifted our focus from the mathematical derivation to the physical analysis, we could start to modify the models in such a way that the dynamics becomes more tractable, and try to understand their physical behaviour.

It could be, in fact, that there exists a whole plethora of interesting background independent theories, which we have not discovered yet because we stuck to quantization schemes that start from classical field theory. What we advocate for instead is to take the new degrees of freedom seriously, and to formulate theories directly with them.

Canonical LQG, gauge-fixing and Lorentz invariance

A central result of canonical LQG is the discreteness of geometric eigenvalues, and it motivates hopes for a UV-finiteness of the theory. The reason for this discreteness lies in the compactness of the gauge group. The latter is, in turn, a result of the adding of the β-dependent term in the action, and the fixing of the gauge in the Hamiltonian formalism (see sec. 1.3.4). In that sense, the gauge-fixing is responsible for having a compact gauge group, and explains why we have discrete eigenvalues, and not continuous eigenvalues which would appear if we used SO(1,3).

The fact that the gauge-fixing is done *before* quantization raises serious doubts about the quantum theory. Due to the time gauge, we have no implementation of local Lorentz transformations, and the whole theory is formulated in terms of a particular Lorentz frame. This could result in an explicit breaking of local Lorentz invariance at the quantum level.

Note that we are not saying that the presence of minimal eigenvalues of area is in contradiction to local Lorentz invariance. Rovelli & Speziale and Livine & Oriti have given arguments [98, 99], which indicate that *per se* the discreteness of geometrical spectra does not contradict continuous Lorentz symmetry, in a similar way as discrete eigenvalues of spin do not stand in contradiction to rotational invariance.

What we claim here, however, is that the gauge-fixing *before* quantization (which has the discrete spectrum as a consequence) could imply a breaking of local Lorentz invariance. Of course, working in a gauge, does not necessarily mean that the result is gauge-dependent, but it could easily happen here, since the gauge-fixing has a crucial influence on the spectra

of gauge-invariant operators. Conversely, this casts also doubts on the discreteness of the spectra, since it is achieved by means of a gauge-fixing that might lead to preferred Lorentz frames.

In that perspective, the following question naturally appears: if the aim is to construct a UV-finite theory, based on discrete eigenvalues of operators, wouldn't it be more honest to use a compact gauge group from the start—i.e. by deforming the Lorentz group in some way—instead of introducing compactness through the back door by gauge-fixing.

Gravity-inspired Lorentz group deformations are, in fact, a very active field of research at present [100, 101, 102]. They are usually motivated by the "naive" argument that an invariant Planck length is incompatible with Lorentz invariance: this reasoning has a loophole, according to [98, 99], since discrete quantum eigenvalues can be in agreement with Lorentz invariance. The counterargument does not explain, however, where the discrete eigenvalues should come from when the full gauge group is non-compact. As we argued, gauge-fixing is a dubious way to arrive at this discreteness. In that sense, it could be still true that a Planck scale cutoff requires a deformation of the Lorentz group.

It would be interesting to investigate if the techniques of LQG could be combined with a deformed Lorentz group and its representation theory. For that, self-dual gravity might be a good starting point, because its Hamiltonian formulation does not involve any gauge-fixing, and one of the reasons that blocked its progress so far was the non-compactness of the gauge group.

5.3 Conclusion

The conceptual approach of canonical LQG and spin foam gravity is promising, since it is adapted to the aim of constructing a background-independent theory. The appearance of abstract networks (or 2-complexes) as the degrees of freedom in the quantum theory seems very appealing, since it reflects the relativity principle of GR: in the absence of a background, the physical information is encoded in the relation between objects, and absolute positions have no meaning. An abstract network of vertices (objects) and edges (relations) appears ideally suited to implement this idea. A further intriguing feature is the use of compact gauge groups that has opened up the possibility of a self-regulation of the theory, based on minimum geometric eigenvalues.

At the same time, LQG faces serious difficulties: in canonical LQG, we have no generally accepted definition of the dynamics, and the gauge-fixing before quantization could imply a breaking of Lorentz invariance. The proposed spin foam models have a specified dynamics, but the spin foam amplitudes are complicated and we do not know if they lead to finite transition amplitudes. Moreover, since these systems differ from anything that appeared in field theory and solid state physics so far, the analysis of their high- and low-energy physics requires the development of completely new methods.

In both loop and spin foam gravity, the challenge is to find a form of the dynamics that is simple enough to be analyzed and understood, and can lead to a physically realistic semiclassical limit: it has to give

- correlations at scales much larger than the fundamental length scale,
- an effective field theory on backgrounds that contains gravity,
- and local Lorentz invariance or a modification of it within experimental limits.

Altogether I consider LQG as a very promising approach, but it might still undergo considerable changes at the kinematic and dynamical level, before it becomes a complete and consistent quantum theory of gravity.

Bibliography

[1] Jean-Michel Drouffe and Jean-Bernard Zuber. Strong coupling and mean field methods in lattice gauge theories. *Phys. Rept.*, 102:1, 1983.

[2] Thomas Thiemann. The LQG string: Loop quantum gravity quantization of string theory. I: Flat target space. *Class. Quant. Grav.*, 23:1923–1970, 2006.

[3] Florian Conrady and Carlo Rovelli. Generalized Schrödinger equation in Euclidean field theory. *Int. J. Mod. Phys.*, A19:4037–4068, 2004.

[4] Florian Conrady. Free vacuum for loop quantum gravity. *Class. Quant. Grav.*, 22: 3261–3293, 2005.

[5] Florian Conrady. Geometric spin foams, Yang-Mills theory and background- independent models. 2005.

[6] Robert Oeckl. Schrödinger's cat and the clock: Lessons for quantum gravity. *Class. Quant. Grav.*, 20:5371–5380, 2003.

[7] Robert Oeckl. A 'general boundary' formulation for quantum mechanics and quantum gravity. *Phys. Lett.*, B575:318–324, 2003.

[8] Robert Oeckl. The general boundary approach to quantum gravity. 2003.

[9] Robert Oeckl. States on timelike hypersurfaces in quantum field theory. *Phys. Lett.*, B622:172–177, 2005.

[10] Florian Conrady, Luisa Doplicher, Robert Oeckl, Carlo Rovelli, and Massimo Testa. Minkowski vacuum in background independent quantum gravity. *Phys. Rev.*, D69: 064019, 2004.

[11] Carlo Rovelli. Quantum gravity. Cambridge, UK: Univ. Pr. (2004) 455 p.

[12] Thomas Thiemann. Introduction to modern canonical quantum general relativity. 2001.

[13] Alejandro Perez. Introduction to loop quantum gravity and spin foams. 2004.

[14] John C. Baez. An introduction to spin foam models of BF theory and quantum gravity. *Lect. Notes Phys.*, 543:25–94, 2000.

[15] Hermann Nicolai, Kasper Peeters, and Marija Zamaklar. Loop quantum gravity: An outside view. *Class. Quant. Grav.*, 22:R193, 2005.

[16] Peter Peldan. Actions for gravity, with generalizations: A review. *Class. Quant. Grav.*, 11:1087–1132, 1994.

[17] Andreas W. Wipf. Hamilton's formalism for systems with constraints. 1993.

[18] M. Henneaux and C. Teitelboim. Quantization of gauge systems. Princeton, USA: Univ. Pr. (1992) 520 p.

[19] H. Nicolai and H. J. Matschull. Aspects of canonical gravity and supergravity. *J. Geom. Phys.*, 11:15–62, 1993.

[20] A. Ashtekar. New variables for classical and quantum gravity. *Phys. Rev. Lett.*, 57: 2244–2247, 1986.

[21] A. Ashtekar. New Hamiltonian formulation of general relativity. *Phys. Rev.*, D36: 1587–1602, 1987.

[22] A. Ashtekar. Lectures on nonperturbative canonical gravity. Singapore, Singapore: World Scientific (1991) 334 p. (Advanced series in astrophysics and cosmology, 6).

[23] J. Fernando Barbero. Real Ashtekar variables for Lorentzian signature space times. *Phys. Rev.*, D51:5507–5510, 1995.

[24] Soren Holst. Barbero's Hamiltonian derived from a generalized Hilbert- Palatini action. *Phys. Rev.*, D53:5966–5969, 1996.

[25] Nuno Barros e Sa. Hamiltonian analysis of general relativity with the Immirzi parameter. *Int. J. Mod. Phys.*, D10:261–272, 2001.

[26] Ted Jacobson and Lee Smolin. Nonperturbative quantum geometries. *Nucl. Phys.*, B299:295, 1988.

[27] Carlo Rovelli and Lee Smolin. Loop space representation of quantum general relativity. *Nucl. Phys.*, B331:80, 1990.

[28] Abhay Ashtekar and C. J. Isham. Representations of the holonomy algebras of gravity and nonabelian gauge theories. *Class. Quant. Grav.*, 9:1433–1468, 1992.

[29] Carlo Rovelli and Lee Smolin. Spin networks and quantum gravity. *Phys. Rev.*, D52: 5743–5759, 1995.

[30] Roger Penrose. Angular momentum: an approach to combinatorial space-time. In Ted Bastin, editor, *Quantum Theory and Beyond*. Cambridge University Press, 1971.

[31] Carlo Rovelli and Lee Smolin. Discreteness of area and volume in quantum gravity. *Nucl. Phys.*, B442:593–622, 1995.

[32] T. Thiemann. Anomaly-free formulation of non-perturbative, four-dimensional Lorentzian quantum gravity. *Phys. Lett.*, B380:257–264, 1996.

[33] T. Thiemann. QSD VI: Quantum Poincare algebra and a quantum positivity of energy theorem for canonical quantum gravity. *Class. Quant. Grav.*, 15:1463–1485, 1998.

[34] Lee Smolin. The classical limit and the form of the Hamiltonian constraint in non-perturbative quantum general relativity. 1996.

[35] Thomas Thiemann. The Phoenix project: Master constraint programme for loop quantum gravity. *Class. Quant. Grav.*, 23:2211–2248, 2006.

[36] Bianca Dittrich and Thomas Thiemann. Testing the master constraint programme for loop quantum gravity. V: Interacting field theories. *Class. Quant. Grav.*, 23:1143–1162, 2006.

[37] Michael P Reisenberger and Carlo Rovelli. Sum over surfaces form of loop quantum gravity. *Phys. Rev.*, D56:3490–3508, 1997.

[38] Karim Noui and Alejandro Perez. Three dimensional loop quantum gravity: Physical scalar product and spin foam models. *Class. Quant. Grav.*, 22:1739–1762, 2005.

[39] Alejandro Perez. Spin foam models for quantum gravity. *Class. Quant. Grav.*, 20:R43, 2003.

[40] Laurent Freidel and David Louapre. Diffeomorphisms and spin foam models. *Nucl. Phys.*, B662:279–298, 2003.

[41] Laurent Freidel and David Louapre. Ponzano-Regge model revisited. II: Equivalence with Chern-Simons. 2004.

[42] Laurent Freidel and David Louapre. Ponzano-Regge model revisited. I: Gauge fixing, observables and interacting spinning particles. *Class. Quant. Grav.*, 21:5685–5726, 2004.

[43] Laurent Freidel and Etera R. Livine. Ponzano-Regge model revisited. III: Feynman diagrams and effective field theory. *Class. Quant. Grav.*, 23:2021–2062, 2006.

[44] Alejandro Perez. Spin foam quantization of SO(4) Plebanski's action. *Adv. Theor. Math. Phys.*, 5:947–968, 2002.

[45] Adam D. Helfer. The stress-energy operator. *Class. Quant. Grav.*, 13:L129–L134, 1996.

[46] M. Atiyah. Topological quantum field theories. *Inst. Hautes Etudes Sci. Publ. Math.*, 68:175, 1989.

[47] S. Tomonaga. On a relativistically invariant formulation of the quantum theory of wave fields. *Prog. Theoret. Phys.*, 1:27–42, 1946.

[48] Julian S. Schwinger. Quantum electrodynamics. I: A covariant formulation. *Phys. Rev.*, 74:1439, 1948.

[49] R. P. Feynman. Space-time approach to nonrelativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, 1948.

[50] Jonathan J. Halliwell and James B. Hartle. Wave functions constructed from an invariant sum over histories satisfy constraints. *Phys. Rev.*, D43:1170–1194, 1991.

[51] Charles G. Torre and Madhavan Varadarajan. Functional evolution of free quantum fields. *Class. Quant. Grav.*, 16:2651–2668, 1999.

[52] L.S. Schulman. *Techniques and Applications of Path Integration.* J. Wiley & Sons, New York, 1981.

[53] Luisa Doplicher. Generalized Tomonaga-Schwinger equation from the Hadamard formula. *Phys. Rev.*, D70:064037, 2004.

[54] Lee Smolin. Quantum gravity with a positive cosmological constant. 2002.

[55] Chopin Soo and Lay Nam Chang. Superspace dynamics and perturbations around 'emptiness'. *Int. J. Mod. Phys.*, D3:529–544, 1994.

[56] A. Mikovic. Quantum gravity vacuum and invariants of embedded spin networks. *Class. Quant. Grav.*, 20:3483–3492, 2003.

[57] A. Mikovic. Flat spacetime vacuum in loop quantum gravity. *Class. Quant. Grav.*, 21: 3909–3922, 2004.

[58] Abhay Ashtekar, Carlo Rovelli, and Lee Smolin. Gravitons and loops. *Phys. Rev.*, D44:1740–1755, 1991.

[59] Madhavan Varadarajan. Gravitons from a loop representation of linearised gravity. *Phys. Rev.*, D66:024017, 2002.

[60] Thomas Thiemann. Gauge field theory coherent states (GCS). I: General properties. *Class. Quant. Grav.*, 18:2025–2064, 2001.

[61] Abhay Ashtekar and Jerzy Lewandowski. Relation between polymer and fock excitations. *Class. Quant. Grav.*, 18:L117–L128, 2001.

[62] Abhay Ashtekar, Carlo Rovelli, and Lee Smolin. Weaving a classical geometry with quantum threads. *Phys. Rev. Lett.*, 69:237–240, 1992.

[63] A. Corichi and J. M. Reyes. A Gaussian weave for kinematical loop quantum gravity. *Int. J. Mod. Phys.*, D10:325–338, 2001.

[64] M. Reuter. Nonperturbative evolution equation for quantum gravity. *Phys. Rev.*, D57: 971–985, 1998.

[65] Roberto Percacci and Daniele Perini. Asymptotic safety of gravity coupled to matter. *Phys. Rev.*, D68:044018, 2003.

[66] Fotini Markopoulou. Coarse graining in spin foam models. *Class. Quant. Grav.*, 20: 777–800, 2003.

[67] Robert Oeckl. Renormalization of discrete models without background. *Nucl. Phys.*, B657:107–138, 2003.

[68] Chopin Soo. Wave function of the universe and Chern-Simons perturbation theory. *Class. Quant. Grav.*, 19:1051–1064, 2002.

[69] Artem Starodubtsev. Topological excitations around the vacuum of quantum gravity. I: The symmetries of the vacuum. 2003.

[70] Lee Smolin. Strings as perturbations of evolving spin-networks. *Nucl. Phys. Proc. Suppl.*, 88:103–113, 2000.

[71] Fotini Markopoulou and Lee Smolin. Causal evolution of spin networks. *Nucl. Phys.*, B508:409–430, 1997.

[72] Fotini Markopoulou and Lee Smolin. Quantum geometry with intrinsic local causality. *Phys. Rev.*, D58:084032, 1998.

[73] Abhay Ashtekar and Joo-Han Lee. Weak field limit of general relativity in terms of new variables: A Hamiltonian framework. *Int. J. Mod. Phys.*, D3:675–694, 1994.

[74] A. Higuchi. Linearized quantum gravity in flat space with toroidal topology. *Class. Quant. Grav.*, 8:2023–2034, 1991.

[75] Thomas Thiemann. Complexifier coherent states for quantum general relativity. *Class. Quant. Grav.*, 23:2063–2118, 2006.

[76] Hanno Sahlmann and Thomas Thiemann. Towards the QFT on curved spacetime limit of QGR. II: A concrete implementation. *Class. Quant. Grav.*, 23:909–954, 2006.

[77] Junichi Iwasaki. Woven geometries: Black holes. 1994.

[78] Michael P. Reisenberger. World sheet formulations of gauge theories and gravity. 1994.

[79] Junichi Iwasaki. A reformulation of the Ponzano-Regge quantum gravity model in terms of surfaces. 1994.

[80] Robert Oeckl and Hendryk Pfeiffer. The dual of pure non-abelian lattice gauge theory as a spin foam model. *Nucl. Phys.*, B598:400–426, 2001.

[81] Michael P. Reisenberger. A lattice worldsheet sum for 4-d Euclidean general relativity. 1997.

[82] John C. Baez. Spin foam models. *Class. Quant. Grav.*, 15:1827–1858, 1998.

[83] Michael P. Reisenberger and Carlo Rovelli. Spacetime as a Feynman diagram: The connection formulation. *Class. Quant. Grav.*, 18:121–140, 2001.

[84] Michael Reisenberger and Carlo Rovelli. Spin foams as Feynman diagrams. 2000.

[85] Martin Bojowald and Alejandro Perez. Spin foam quantization and anomalies. 2003.

[86] Jose A. Zapata. Continuum spin foam model for 3d gravity. *J. Math. Phys.*, 43: 5612–5623, 2002.

[87] Ivan K. Kostov. U(N) gauge theory and lattice strings. *Nucl. Phys.*, B415:29–70, 1994.

[88] John C. Baez. Spin network states in gauge theory. *Adv. Math.*, 117:253–272, 1996.

[89] Florian Girelli, Robert Oeckl, and Alejandro Perez. Spin foam diagrammatics and topological invariance. *Class. Quant. Grav.*, 19:1093–1108, 2002.

[90] I. Montvay and G. Münster. Quantum fields on a lattice. Cambridge, UK: Univ. Pr. (1994) 491 p. (Cambridge monographs on mathematical physics).

[91] Roger Dearnaley. Nambu-Goto string theory on a space-time lattice: Plaquette random surfaces revisited. *Int. J. Mod. Phys.*, A8:3339–3358, 1993.

[92] John W. Barrett and Louis Crane. Relativistic spin networks and quantum gravity. *J. Math. Phys.*, 39:3296–3302, 1998.

[93] Roberto De Pietri, Laurent Freidel, Kirill Krasnov, and Carlo Rovelli. Barrett-Crane model from a Boulatov-Ooguri field theory over a homogeneous space. *Nucl. Phys.*, B574:785–806, 2000.

[94] Alejandro Perez and Carlo Rovelli. A spin foam model without bubble divergences. *Nucl. Phys.*, B599:255–282, 2001.

[95] Gernot Münster. Critical phenomena, strings, and interfaces. 1997.

[96] Juan M. Maldacena. The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.*, 2:231–252, 1998.

[97] Leonardo Modesto and Carlo Rovelli. Particle scattering in loop quantum gravity. *Phys. Rev. Lett.*, 95:191301, 2005.

[98] Carlo Rovelli and Simone Speziale. Reconcile Planck-scale discreteness and the Lorentz-Fitzgerald contraction. *Phys. Rev.*, D67:064019, 2003.

[99] Etera R. Livine and Daniele Oriti. About Lorentz invariance in a discrete quantum setting. *JHEP*, 06:050, 2004.

[100] Giovanni Amelino-Camelia. Relativity in space-times with short-distance structure governed by an observer-independent (Planckian) length scale. *Int. J. Mod. Phys.*, D11:35–60, 2002.

[101] Jerzy Kowalski-Glikman. Introduction to doubly special relativity. *Lect. Notes Phys.*, 669:131–159, 2005.

[102] Joao Magueijo and Lee Smolin. Generalized Lorentz invariance with an invariant energy scale. *Phys. Rev.*, D67:044017, 2003.

[103] Hendryk Pfeiffer. Dual variables and a connection picture for the Euclidean Barrett-Crane model. *Class. Quant. Grav.*, 19:1109–1138, 2002.

Appendix A

Local Form of Hamilton-Jacobi and Schrödinger Equation

In the text we have presented the generalized Hamilton-Jacobi and Schrödinger equation in an integral form. They can be also expressed locally, and below we explain how the two representations are related. The local notation is used in [11] and [10].

Both the Hamilton function S and the propagator W depend on the volume V. The latter is enclosed by the boundary Σ. Consider a parametrization of Σ, i.e. a map

$$x : P \to \Sigma\,,\ \ \tau \mapsto x(\tau)$$

from a $(d-1)$-dimensional manifold P to Σ. Provided it is clear on which "side" of Σ the volume lies, one can view S and W as functions of Σ, or equivalently, as functionals of the parametrizing map $\tau \mapsto x(\tau)$. The other variable of S and W is the boundary field $\varphi : \Sigma \to \mathbb{R}$: we may replace it by its pull-back $\tilde{\varphi}$ to P, so that S and W are completely expressed in terms of quantities on the parameter manifold P:

$$\begin{aligned}&\tilde{\varphi}(\tau) = \varphi(x(\tau))\,,\ \ \tau \in P\,,\\ &S = S[\tilde{\varphi}(\tau), x(\tau)]\,,\quad W = W[\tilde{\varphi}(\tau), x(\tau)]\,.\end{aligned}$$

In section 2.3, we defined the deformation derivative L_N which acts by infinitesimal diffeomorphisms and pull-forwards of V and φ respectively. A moment's reflection shows that in the new notation the same effect is achieved by applying a variation

$$\delta x(\tau) = sN(x(\tau))$$

to the function $x(\tau)$ while leaving $\tilde{\varphi}(\tau)$ fixed. Therefore,

$$L_N \equiv \int_P \mathrm{d}^{d-1}\tau\, N^\mu(x(\tau)) \frac{\delta}{\delta x^\mu(\tau)}\,. \tag{A.1}$$

Our explicit result for the Hamilton-Jacobi equation (p. 46, eq. (2.22)) can be rewritten as

$$\begin{aligned}&L_N S[\varphi, V]\\ &\quad= \int_P \mathrm{d}^{d-1}\tau\, N^\mu(x(\tau)) \Big\{ n_\mu(x(\tau)) \left[-\frac{1}{2}\left(\frac{\delta S}{\delta\varphi(\tau)}\right)^2 + \frac{1}{2}(\nabla\phi(\tau))^2 + \frac{1}{2}m^2\phi^2(\tau) + U(\phi(\tau)) \right]\\ &\qquad - \nabla_\mu\varphi(\tau)\frac{\delta S}{\delta\varphi(\tau)} \Big\}\,,\end{aligned} \tag{A.2}$$

where on the right-hand side S is a functional of the new variables. Comparison with (A.1) gives the equation

$$\frac{\delta S}{\delta x^\mu(\tau)} \;=\; n_\mu(x(\tau)) \left[-\frac{1}{2}\left(\frac{\delta S}{\delta\varphi(\tau)}\right)^2 + \frac{1}{2}(\nabla\phi(\tau))^2 + \frac{1}{2}m^2\phi^2(\tau) + U(\phi(\tau)) \right] - \nabla_\mu\varphi(\tau)\frac{\delta S}{\delta\varphi(\tau)}\,. \tag{A.3}$$

It describes how S behaves under local variations of the boundary Σ. By the same reasoning, we arrive at a local Schrödinger equation for the kernel W:

$$\frac{\delta W}{\delta x^{\mu}(\tau)} = n_{\mu}(x(\tau)) \left[-\frac{\hbar^2}{2} \frac{\delta^2 W}{\delta \varphi(\tau)^2} + \frac{1}{2} (\nabla \phi(\tau))^2 + \frac{1}{2} m^2 \phi^2(\tau) \right] - \hbar \nabla_{\mu} \varphi(\tau) \frac{\delta W}{\delta \varphi(\tau)} \,. \tag{A.4}$$

Appendix B

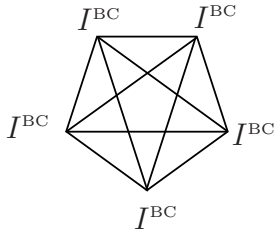
Topological invariance of Barrett-Crane amplitudes

In the original paper on their model, Barrett and Crane fixed only the vertex amplitude, while the amplitudes for edges and faces remained unspecified. Since then various versions of the Barrett-Crane model have been proposed that differ in the choice of edge and face amplitudes. We show below that the model B in [93] is the version which satisfies our requirement of topological invariance.

We start from the conventional formulation where spin foams are viewed as oriented 2-complexes κ with intertwiner labels on edges and irrep labels on faces. Representations are allowed to be trivial. The choice of κ is fixed to a single complex—the dual 2-skeleton of a triangulation. Then, the partition function is given by

$$Z = \left(\prod_{f\in\kappa}\sum_{\rho_f^{\mathsf{bal}}}\right)\left(\prod_{f\in\kappa}\dim V_{\rho_f^{\mathsf{bal}}}\right)\left(\prod_{v\in\kappa}A_v\right) \tag{B.1}$$

where the sum runs over all possible ways to assign balanced representations of SO(4) to faces. The vertex amplitude results from a contraction of normalized Barrett-Crane intertwiners:

$A_v =$ 

(B.2)

Each vertex of the pentagon corresponds to an edge of the complex, and each edge carries a balanced representation from a face of the complex.

Let us translate this to the geometric language with unbranched components and branching lines. Within an unbranched component, faces are non-trivially labelled and arranged in a surface. Faces that "go off" have the trivial label. In the vertex diagram (B.2), this corresponds to having an open or closed line of non-trivial irreps, while other edges are trivial (see Fig. B.1).

To evalute the diagram, we use the definition of the Barrett-Crane intertwiners (see e.g. [103]). Recall that a balanced representation of SO(4) is isomorphic to a tensor product $j \otimes j^*$ where j is a representation of SU(2). A BC-intertwiner between balanced irreps can be expressed as a Haar intertwiner of the associated SU(2) irreps. In the 2-valent case, we

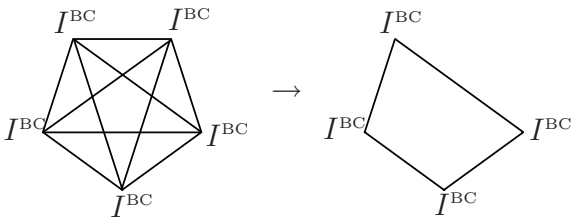
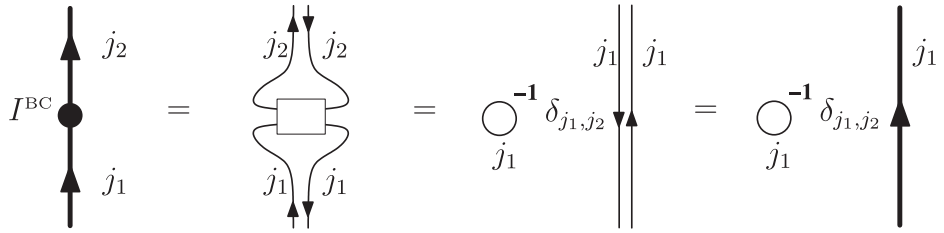


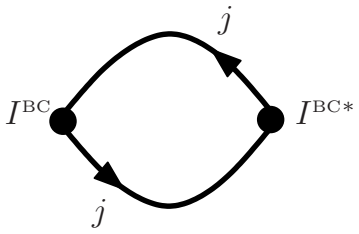
Figure B.1: Vertex amplitude in an unbranched component.

get



(B.3)

The thick lines represent contractions w.r.t. balanced representations, while thin lines stand for the SU(2) irreps. We see that (B.3) is indeed a *normalized* intertwiner, since



$$= \quad 1 \tag{B.4}$$

The identity (B.3) tells us that in Fig. B.1 all representations along a closed line have to be the same, and that open lines would give zero. As in section 4.4, we conclude that unbranched components have to be single-colored and cannot have open ends. For a closed line, we evaluate the contraction of 2-valent BC-intertwiners and obtain $(\dim V_j)^{-1}$ for each intertwiner and $(\dim V_j)^{+2}$ from the contraction. The intertwiner on a lattic edge appears in two vertex amplitudes, so an edge contributes a factor

$$(\dim V_j)^{-2}\,. \tag{B.5}$$

The contraction gives

$$(\dim V_j)^{+2} \tag{B.6}$$

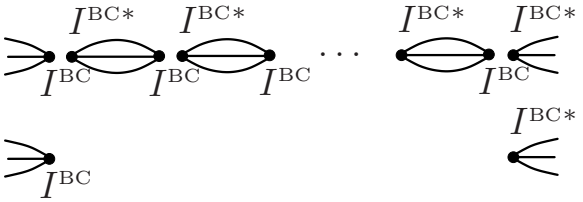
for each vertex, and the face factor in (B.1) yields the dimension of the balanced representation, which is

$$(\dim V_j)^{+2}\,. \tag{B.7}$$

For an unbranched component $\mathcal{F}_i$ with color ρ_i^{bal}, the factors add up to

$$\left(\dim V_{\rho_i^{\mathsf{bal}}}\right)^{\chi(\mathcal{F}_i)}\,. \tag{B.8}$$

It remains to check that contributions from the branching graph are topological. Vertices on the branching line contribute a sequence of factors



(B.9)

Since the BC-intertwiners of the model are normalized, the sequence collapses to a single pair of BC-intertwiners, each of which is associated to a vertex of the branching graph. Thus, we arrive at the geometric form of the spin foam sum that was already stated in equation (4.57).

Acknowledgements

The three years of my Ph.D. were divided into two phases: the time at the Università di Roma "La Sapienza", and the stay at the Centre de Physique Théorique Marseille. I thank the people of the physics department of Rome, who welcomed me so warmly, and did their best to integrate me into their group. Likewise, I thank the CPT Marseille which provided an excellent environment for the second part of my Ph.D.

I also thank the people at the Perimeter Insitute Waterloo and the AEI Potsdam for their hospitality during my visits.

The financial support for the project came from the German Academic Exchange Service (DAAD) and the Daimler-Benz foundation. The AEI Potsdam provided travel support.

For many helpful and interesting discussions I am grateful to the following people: Daniele Oriti, Massimo Testa, Daniele Perini, Pierre Martinetti, Abhay Ashtekar, Lee Smolin, Laurent Freidel, Robert Oeckl and Alejandro Perez.

Finally, I want to thank my two supervisors: Thomas Thiemann and Carlo Rovelli. Thomas was always there to give me encouragement or technical advice when we met in Potsdam or at conferences. I thank Carlo for his willingness to be my "full time" supervisor, for his pragmatism and physical intuition, his courage to go new ways, and for giving me the freedom to follow own ideas when I wanted to do that.

Selbständigkeitserklärung

Ich versichere, daß ich diese Arbeit selbständig verfaßt habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.